KATEDRA FYZIKÁLNÍ CHEMIE
UNIVERZITY PALACKÉHO V OLOMOUCI

INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE

elixir CZECH REPUBLIC

8th Advanced *in silico* Drug Design

KFC/ADD

Structure source

Karel Berka

UP Olomouc, 27.1.-31.1. 2025

ÚOCHB AV ČR IOCB PRAGUE

RECEPTOR.AI

Deep MedChem

Bar-Ilan University אוניברסיטת בר-אילן

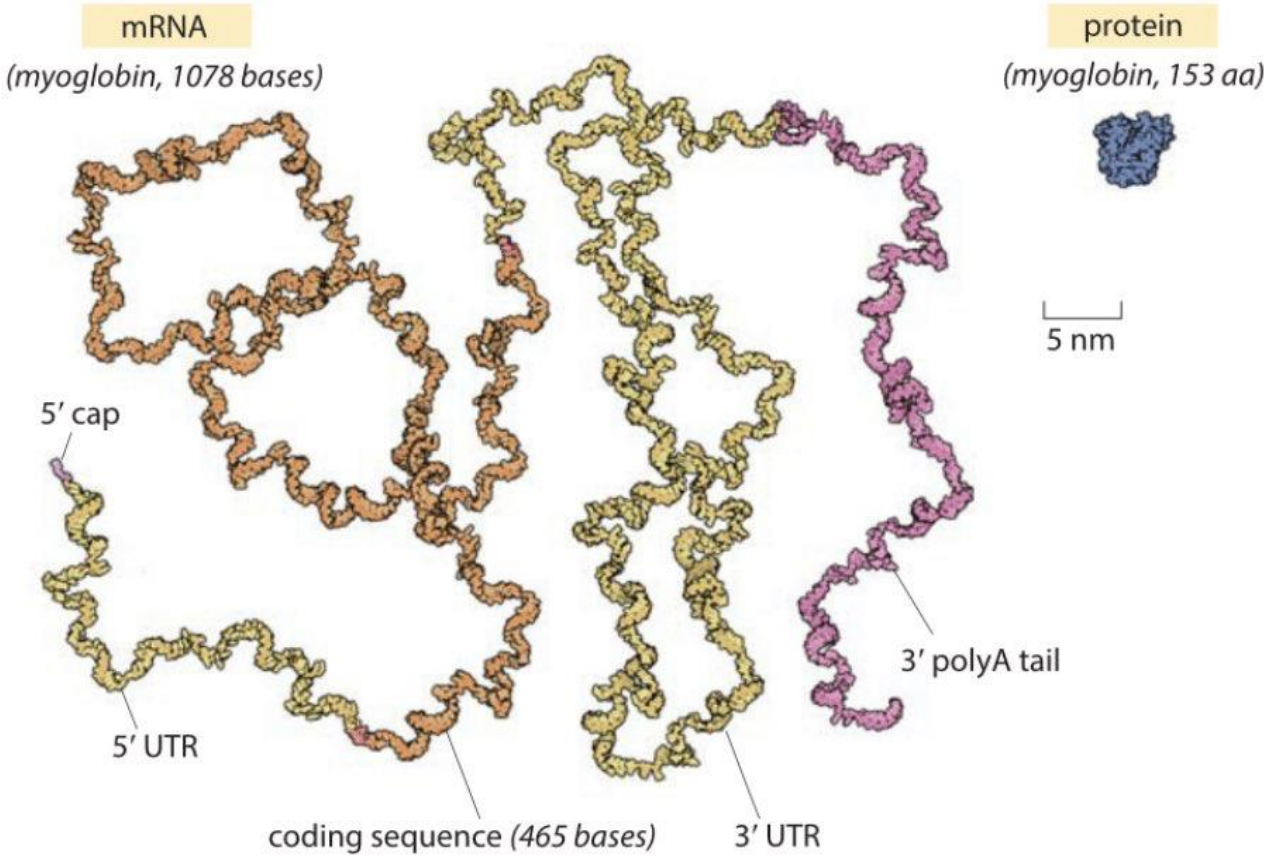
Carnegie Mellon University

Outline

- Sources of structures – ligands and macromolecules
- Macromolecular structure for function
- PDB database and files
- Methods how to get structure – XRAY, NMR, EM, MS, AF
- No structure case - Disorder
- Aggregated view PDBe-KB

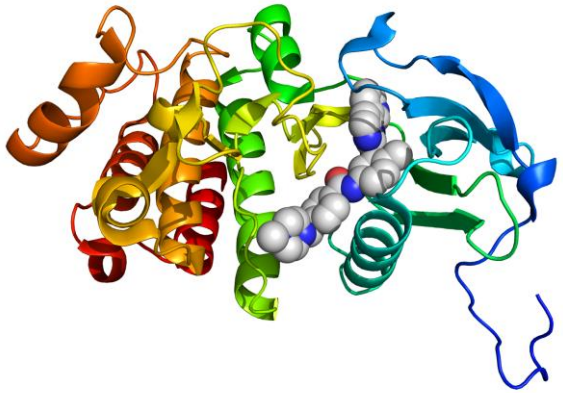
Drug design related structural databases

- Ligands (small molecules)
  - drugbank.ca – comprehensive drug&target info
  - ebi.ac.uk/chembl - bioactive molecules
  - pubchem.ncbi.nlm.nih.gov – free chemical info
  - zinc.docking.org – commercially available compounds for VS
- Targets (proteins/nucleic acids)
  - ebi.ac.uk/pdbe or www.rcsb.org – **macromolecular structures**

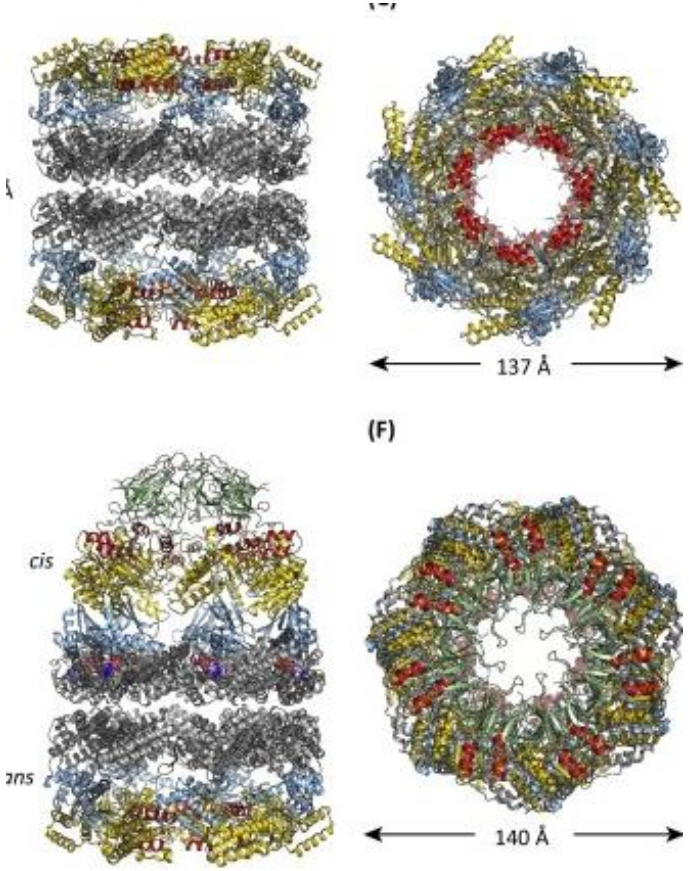


MACROMOLECULAR STRUCTURE

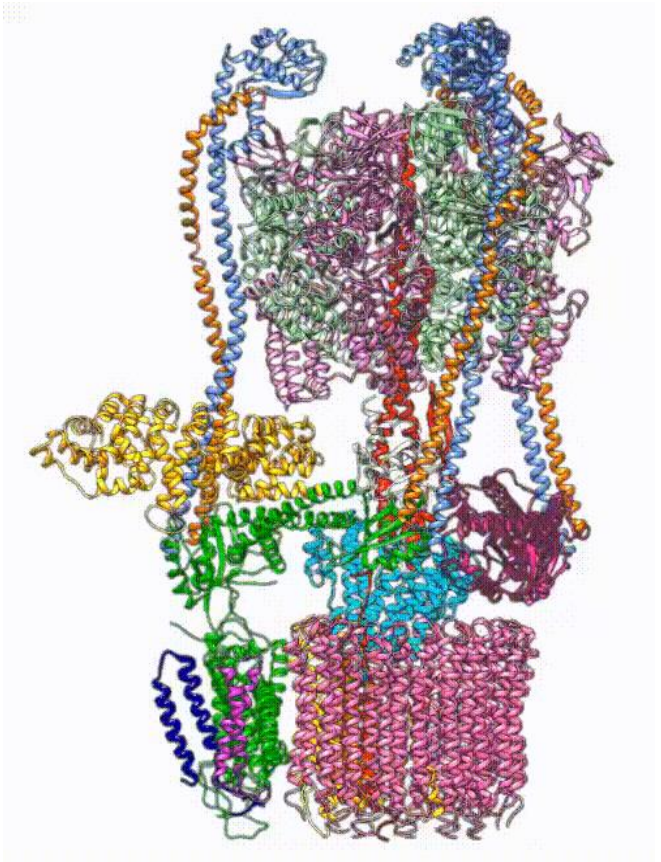
Knowing structure helps to understand the function



imatinib + Abl kinase
PDBID: 2HYY

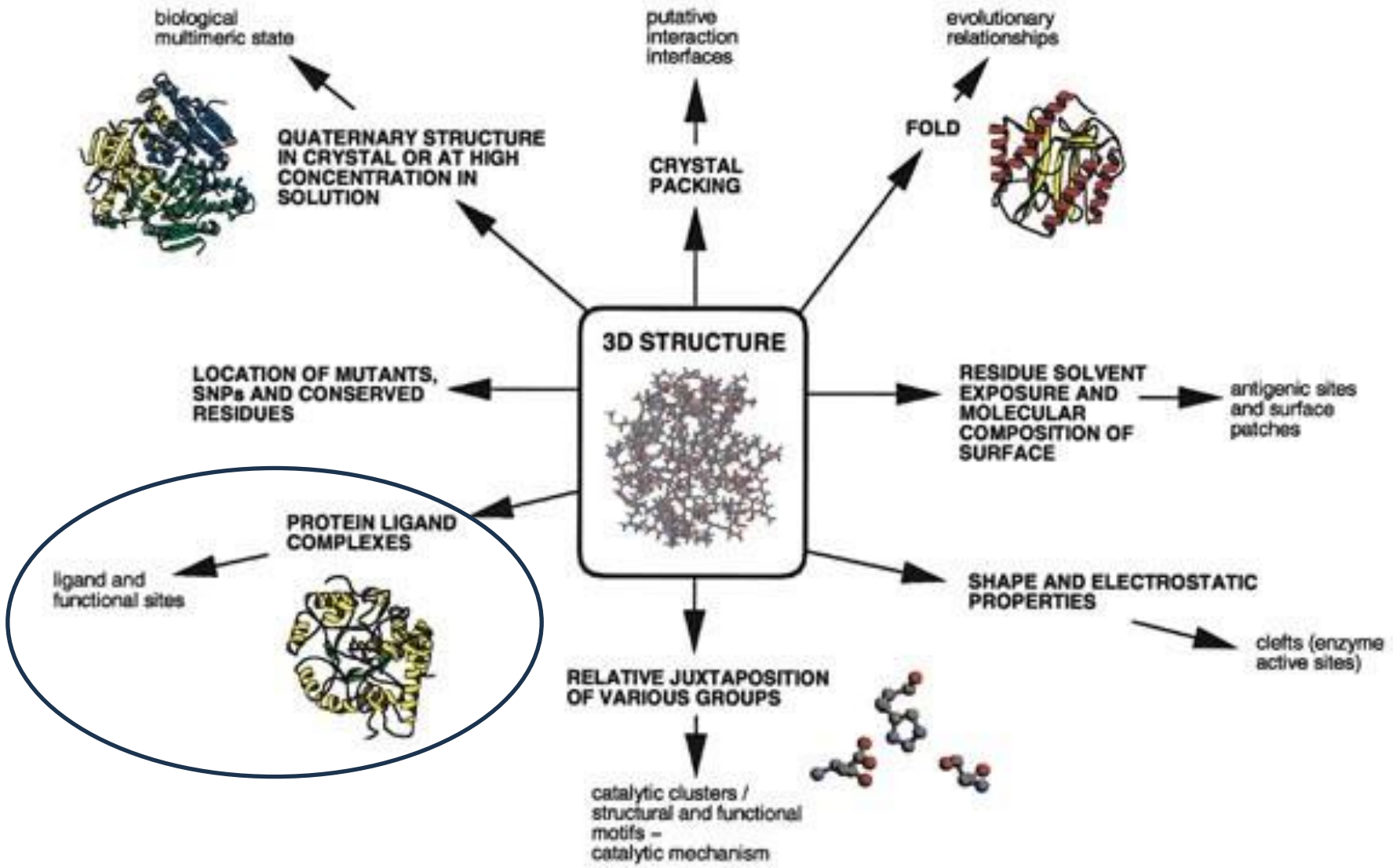


Hayer-Hartl et al., 2015



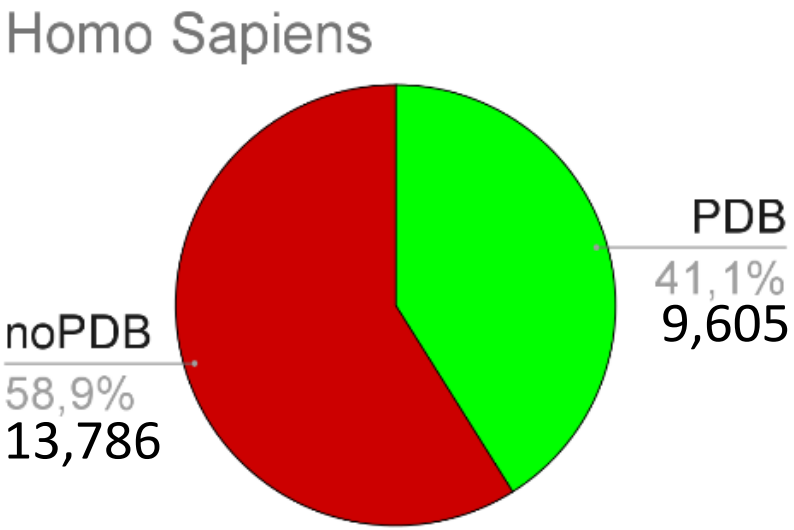
Soung-Hun Roh *et al.* Cryo-EM and MD infer water-mediated proton transport and autoinhibition mechanisms of V_o complex. *Sci.Adv.* **6** ,eabb9605 (2020). DOI:10.1126/sciadv.abb9605

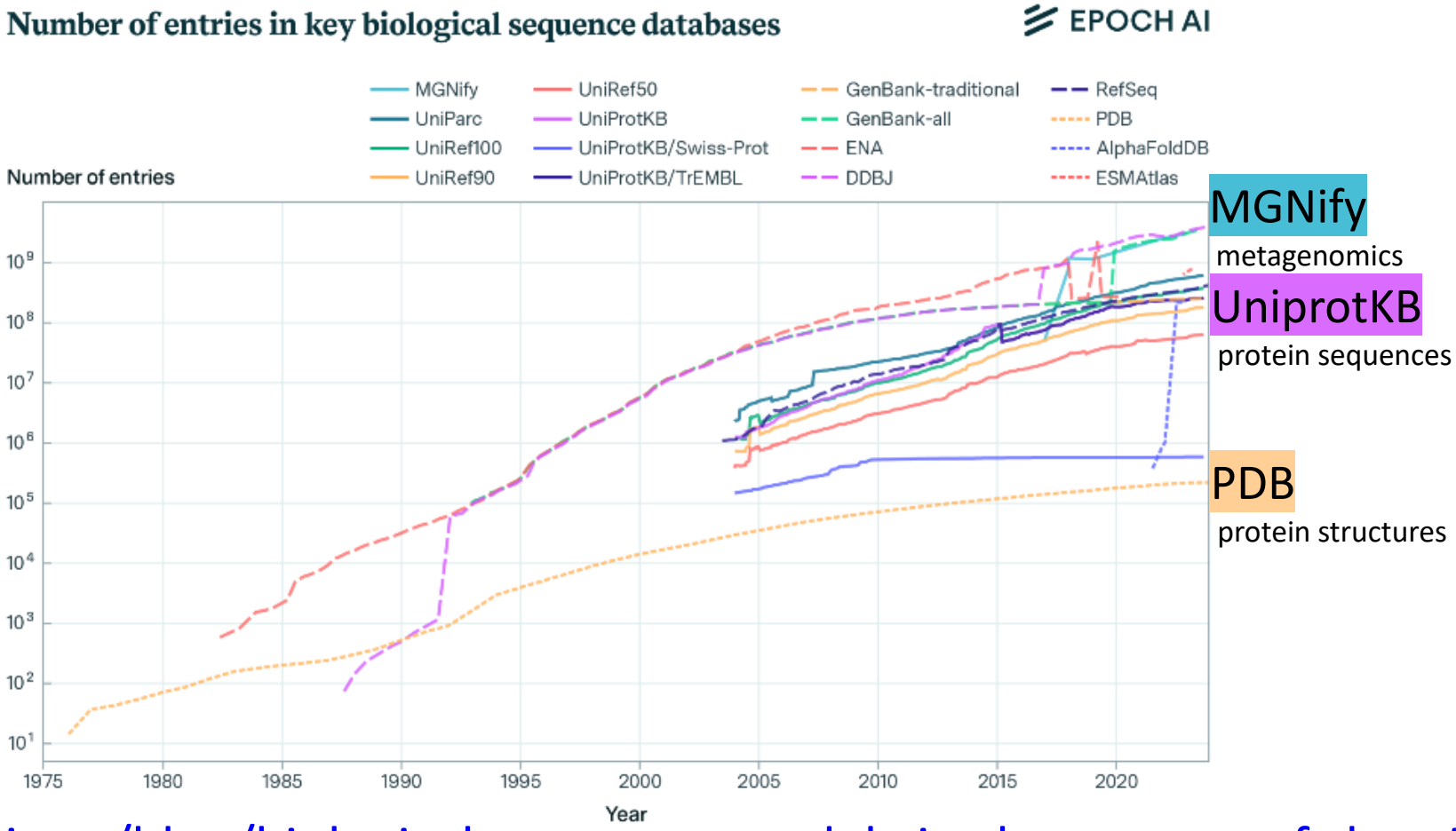
Structure applications for function



Thornton, J., Todd, A., Milburn, D. *et al.* From structure to function: Approaches and limitations. *Nat Struct Mol Biol* **7** (Suppl 11), 991–994 (2000). https://doi.org/10.1038/80784

Gathering of structures is expensive…





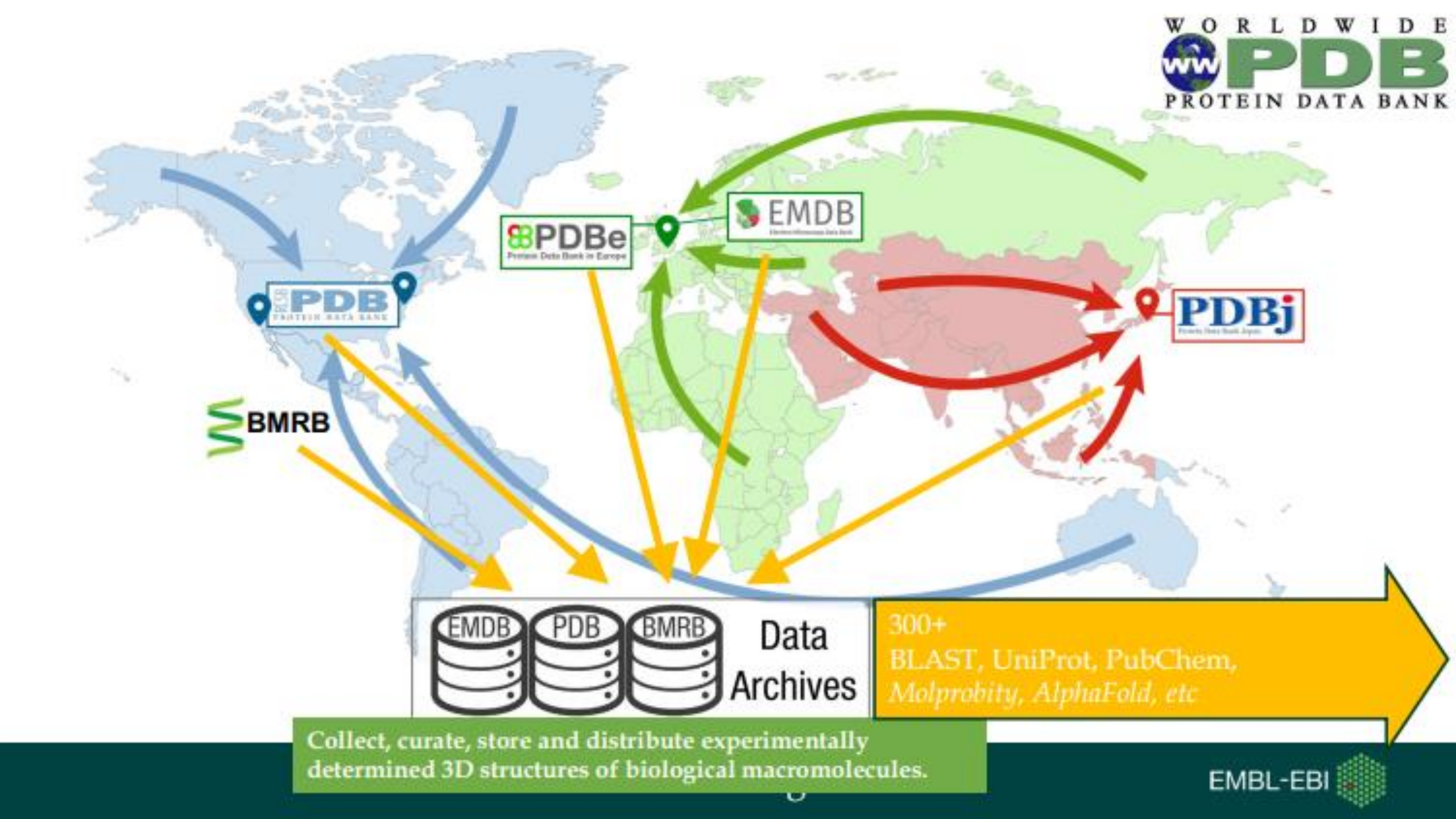
https://epochai.org/blog/biological-sequence-models-in-the-context-of-the-ai-directives

Rozdíl mezi počtem experimentálních struktur a sekvencemi se v čase zvětšuje.

Zjišťování struktur se nezlevňuje tak moc jako sekvenace.

Database of protein structures

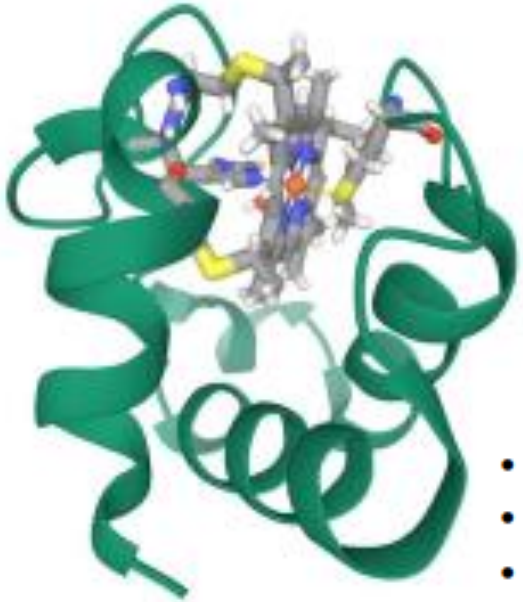
PDB


WORLDWIDE PDB
PROTEIN DATA BANK
PDB
PDBe
Protein Data Bank in Europe
EMDB
PDBj
BMRB
EMDB
PDB
BMRB
Data Archives
300+
BLAST, UniProt, PubChem,
Molprobity, AlphaFold, etc
Collect, curate, store and distribute experimentally determined 3D structures of biological macromolecules.
EMBL-EBI

The Worldwide Protein Data Bank (wwPDB)

EMBL-EBI

A “PDB code” refers to a structure



Described in a paper
(or maybe not published)

- Unique code, currently 4 characters
- Identifies the data within the PDB archive
- Always starts with a number

- *e.g.* 2ins, 4xyz, 2f48
  → pdb_00002ins, pdb_00004xyz, pdb_00002f48

Deposit in the PDB

PDB code

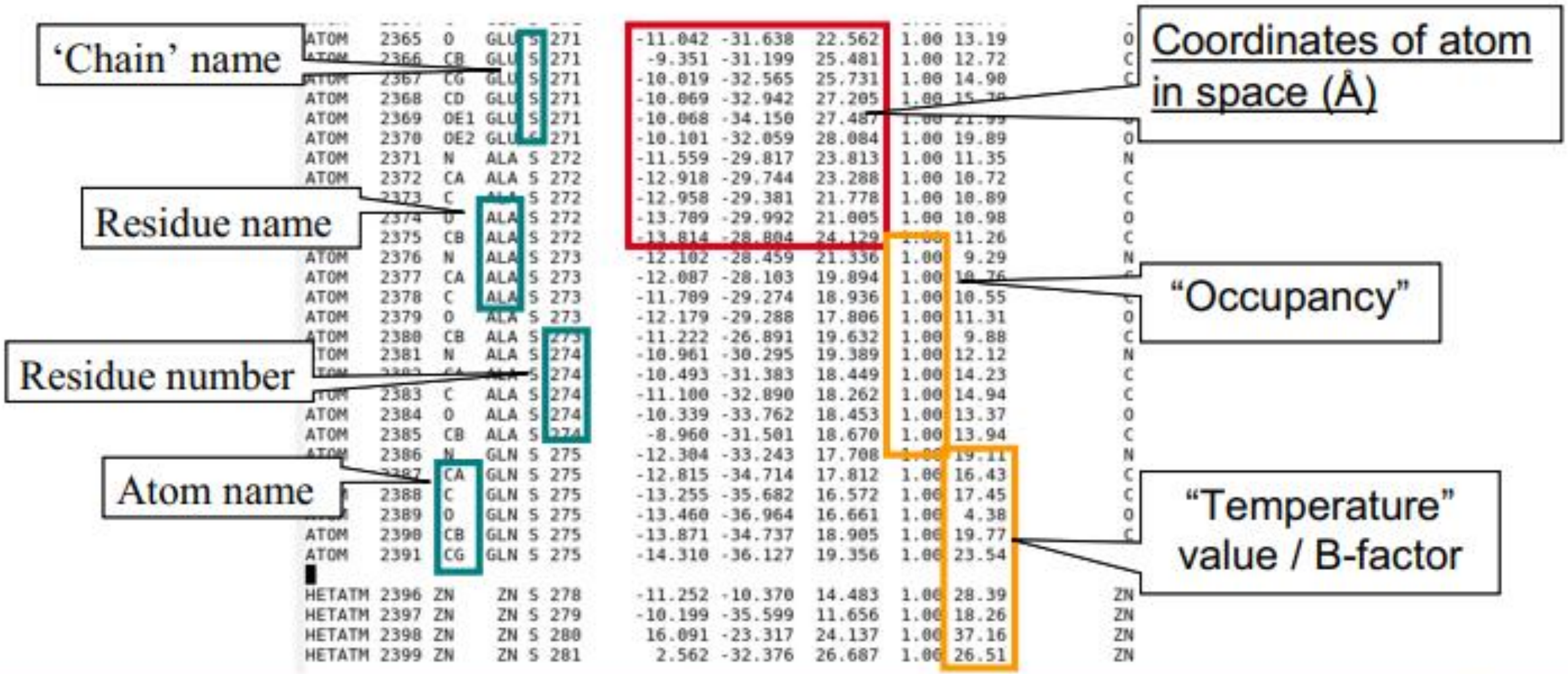
Referenced in the paper


nature

Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB

https://www.wwpdb.org/documentation/new-format-for-pdb-ids

What does a PDB file look like?
A text file with fixed column width - Card legacy



mmCIF is the ‘master format’
It’s still a text file!
More (modern) computer readable

The fastest open-source mmCIF parser:
GEMMI (https://gemmi.readthedocs.io/en/latest/)

```
loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.pdbx_PDB_ins_code
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.Cartn_x_esd
_atom_site.Cartn_y_esd
_atom_site.Cartn_z_esd
_atom_site.occupancy_esd
_atom_site.B_iso_or_equiv_esd
_atom_site.pdbx_formal_charge
_atom_site.auth_seq_id
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_atom_id
_atom_site.pdbx_PDB_model_num
ATOM 1    N N    . MET A 1 1  ? 4.940   -11.534 16.748 1.00 0.00 ? ? ? ? ? ? 1  MET A N    1
ATOM 2    C CA   . MET A 1 1  ? 4.627   -10.082 16.703 1.00 0.00 ? ? ? ? ? ? 1  MET A CA   1
ATOM 3    C C    . MET A 1 1  ? 4.375   -9.616  15.273 1.00 0.00 ? ? ? ? ? ? 1  MET A C    1
ATOM 4    O O    . MET A 1 1  ? 4.604   -10.360 14.318 1.00 0.00 ? ? ? ? ? ? 1  MET A O    1
ATOM 5    C CB   . MET A 1 1  ? 5.800   -9.312  17.310 1.00 0.00 ? ? ? ? ? ? 1  MET A CB   1
ATOM 6    C CG   . MET A 1 1  ? 7.097   -9.468  16.535 1.00 0.00 ? ? ? ? ? ? 1  MET A CG   1
ATOM 7    S SD   . MET A 1 1  ? 8.437   -8.471  17.215 1.00 0.00 ? ? ? ? ? ? 1  MET A SD   1
ATOM 8    C CE   . MET A 1 1  ? 8.819   -9.388  18.706 1.00 0.00 ? ? ? ? ? ? 1  MET A CE   1
ATOM 9    H H1   . MET A 1 1  ? 5.323   -11.746 17.691 1.00 0.00 ? ? ? ? ? ? 1  MET A H1   1
ATOM 10   H H2   . MET A 1 1  ? 5.641   -11.730 16.086 1.00 0.00 ? ? ? ? ? ? 1  MET A H2   1
ATOM 11   H H3   . MET A 1 1  ? 4.856   -12.055 16.578 1.00 0.00 ? ? ? ? ? ? 1  MET A H3   1
ATOM 12   H HA   . MET A 1 1  ? 3.740   -9.904  17.292 1.00 0.00 ? ? ? ? ? ? 1  MET A HA   1
ATOM 13   H HB2  . MET A 1 1  ? 5.548   -8.261  17.340 1.00 0.00 ? ? ? ? ? ? 1  MET A HB2  1
```

Coordinates of atom in space (Å)

Residue name

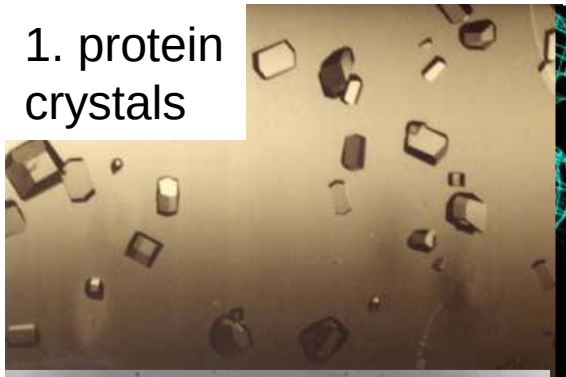
“Occupancy”

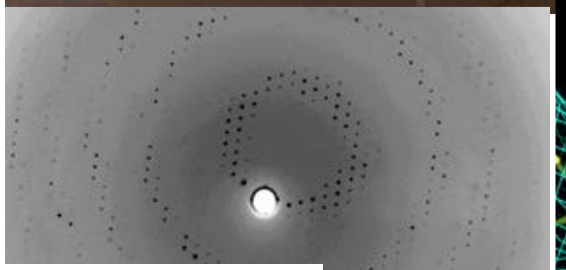
HOW TO GET STRUCTURE OF MACROMOLECULES

Methods

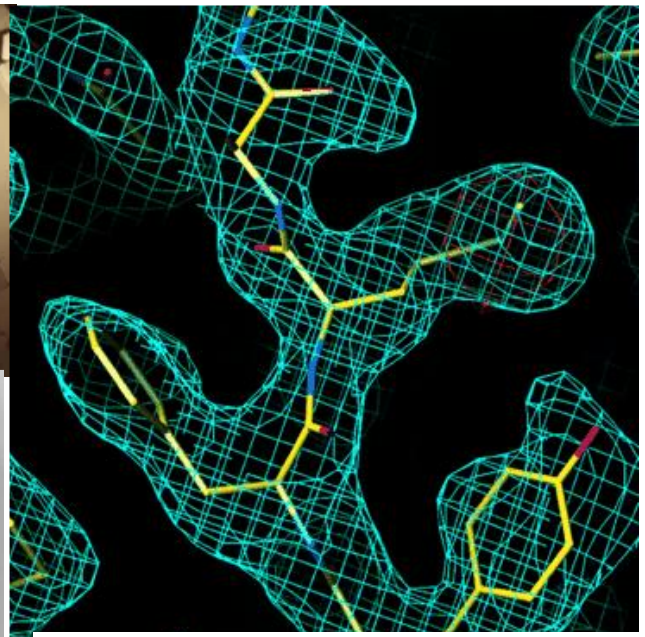
- RTG
  - xyz coordinates
  - inner electron shells
  - crystalization, atomic resolution,
  - interpretation of intermolecular interactions
- EM
  - electron shell
  - low resolution
  - large complexes
- NMR
  - torsion angles and distances
  - dynamical information available
  - MD model
- MS
  - distances
  - molecular weights
  - solvent accesibility
- Modelling
  - AlphaFold

1. protein crystals





2. diffraction experiment

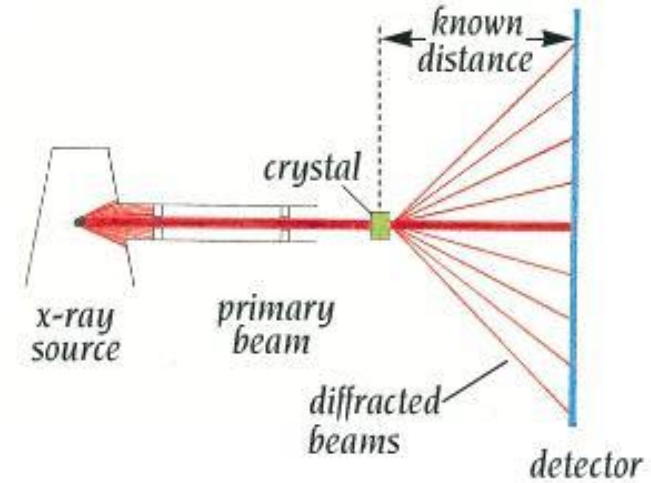


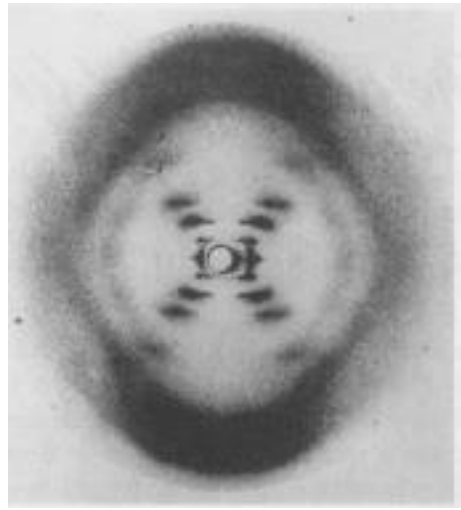
3. map of electron density
4. model fit

https://www.wwpdb.org/

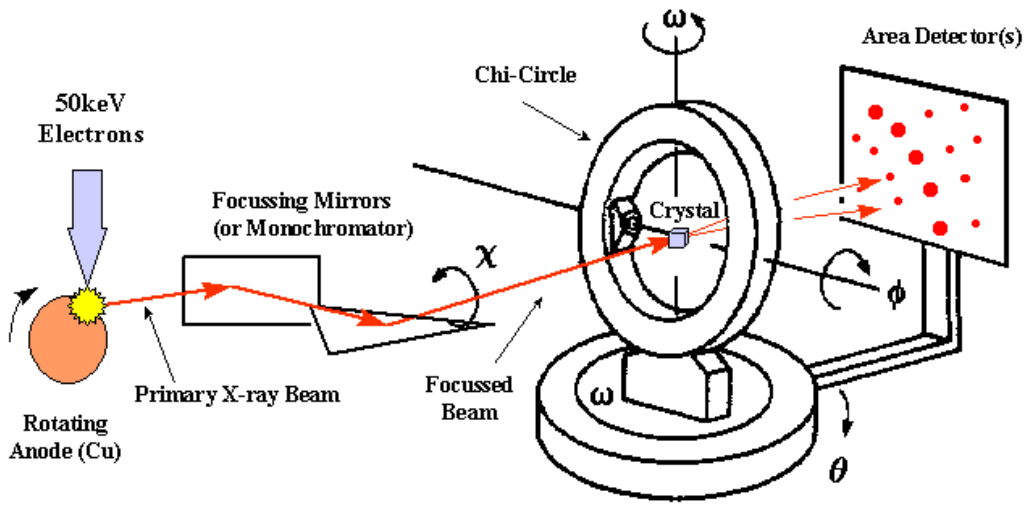
XRAY CRYSTALLOGRAPHY

X-ray Diffraction

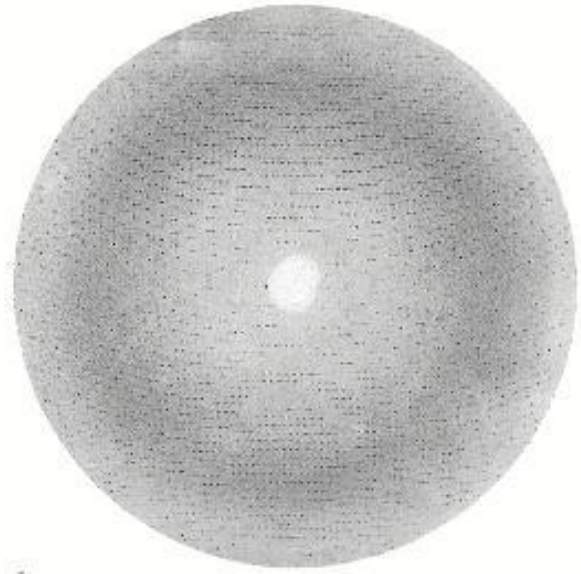




Rosalind Franklin & Raymond Gosling
Nature 171 (1953)

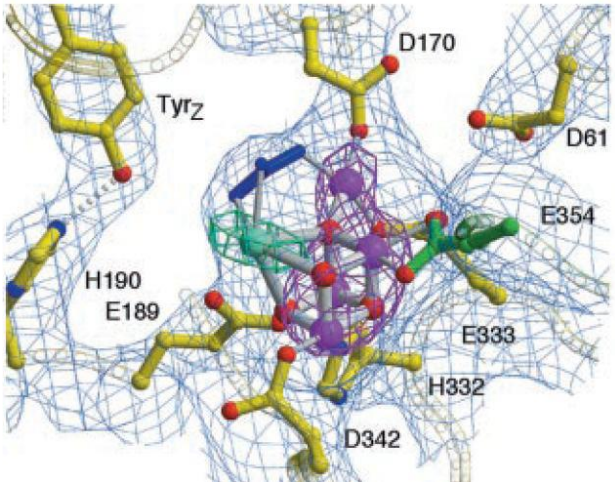


4-Circle Goniometer (Eulerian or Kappa Geometry)

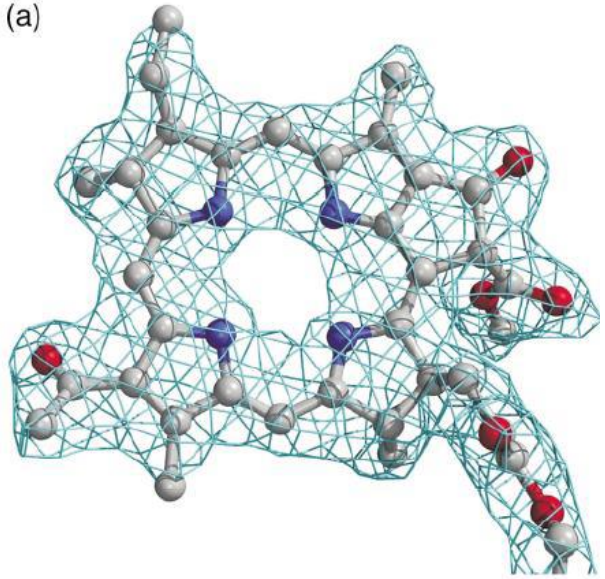


Resolution (R)

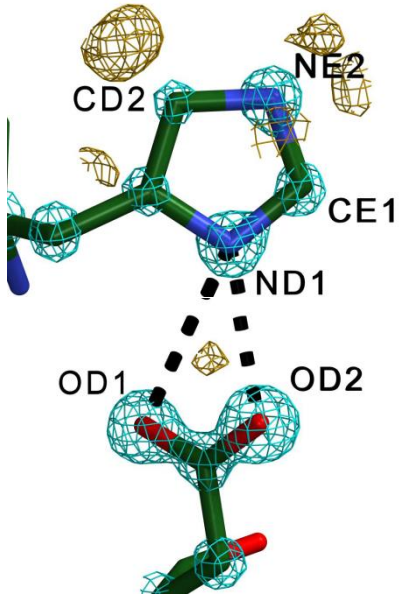
- in Å,
- Distance for distinguishing of two points. Details should be distinguishable at 0.7·R.
- better R – easier model building!
- (more refflections – better signal-to-noise ratio)



3.5 Å map of photosystem II



2.3 Å map of photosynthetic reaction centre



0.95 Å map of elastase

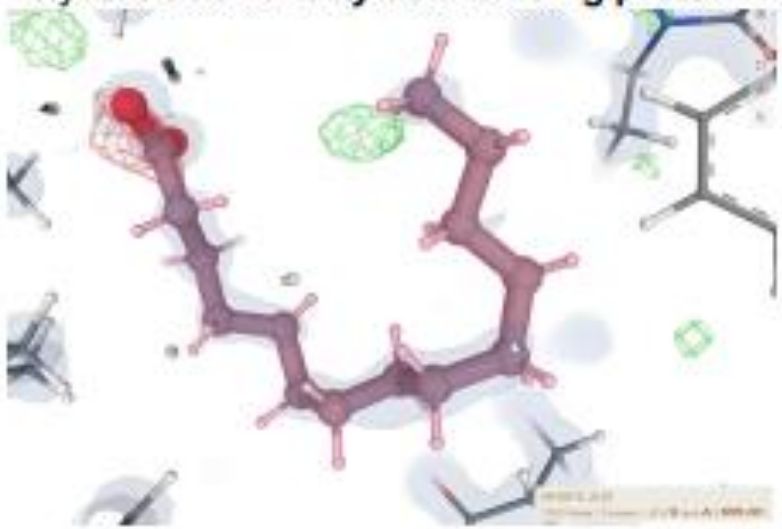
Viewing electron density in Mol*

- To initiate electron maps display:
  clicking on ligand or protein amino acid
- Regular map (blue) '2Fo-Fc' electron density map
  → *should surround atoms*
- Negative and positive density
  → *highlights extra and missing atoms, respectively*

Myristic acid in **Fatty acid-binding protein 4**

PDB ID: 7G0X

Myristic acid in **Fatty acid-binding protein 4**



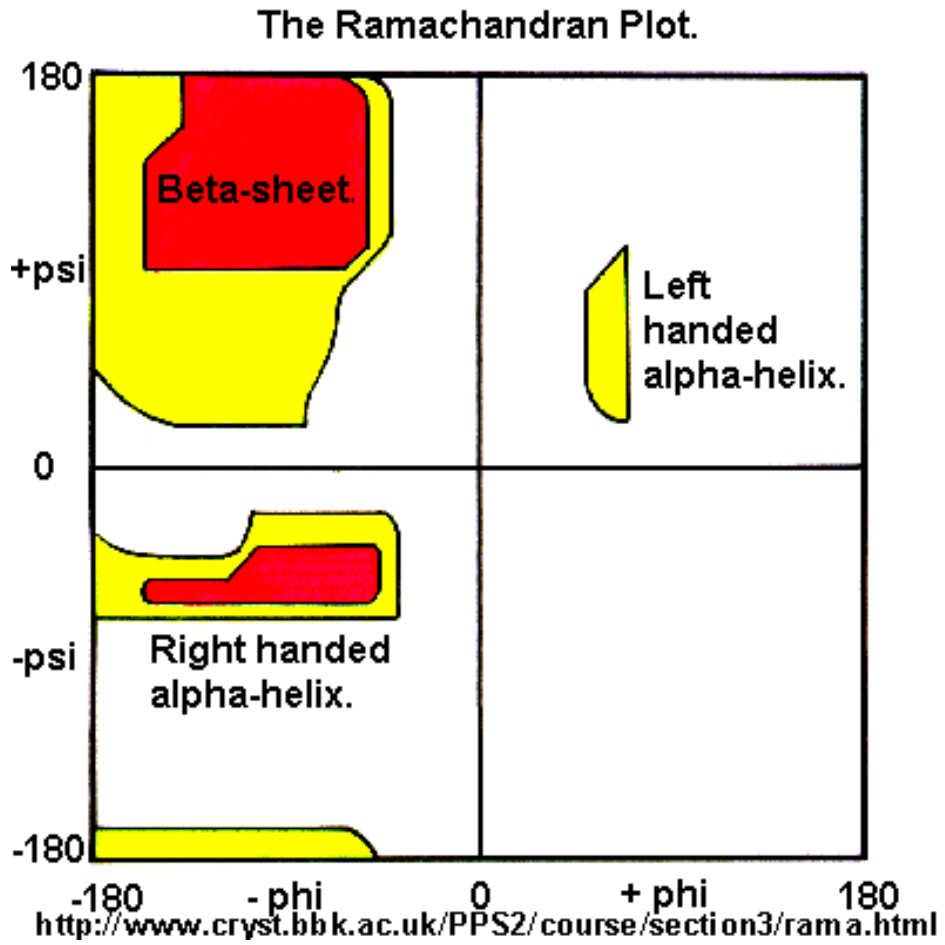
PDB ID: 7FZK

Validation

- R_{free} (Brünger, 1992)
  - test set (~5-10%) of reflection left out from fit for testing
- Stereochemistry

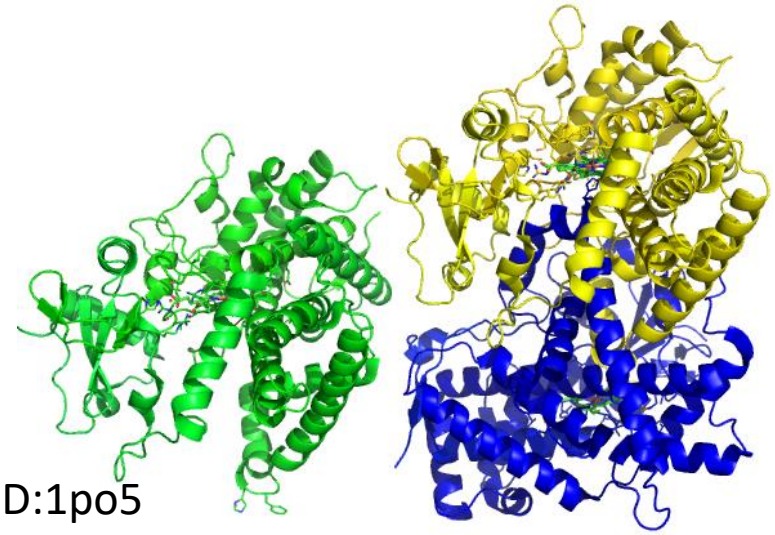
Ramachandran diagram

  - WHATIF
  - MOLPROBITY
- Bad contacts
  - stérické problémy ve struktuře

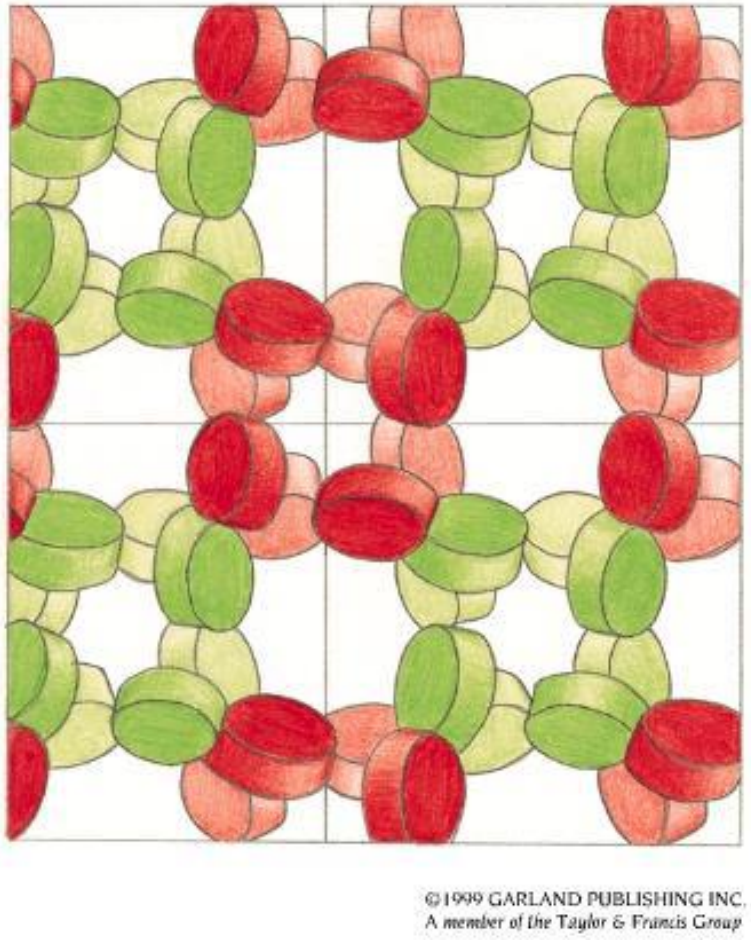


Crystal Contacts

- Protein crystals contain a lot of solvent
- Molecular contacts within crystal do not have usually (well **not always**) effect on protein structure



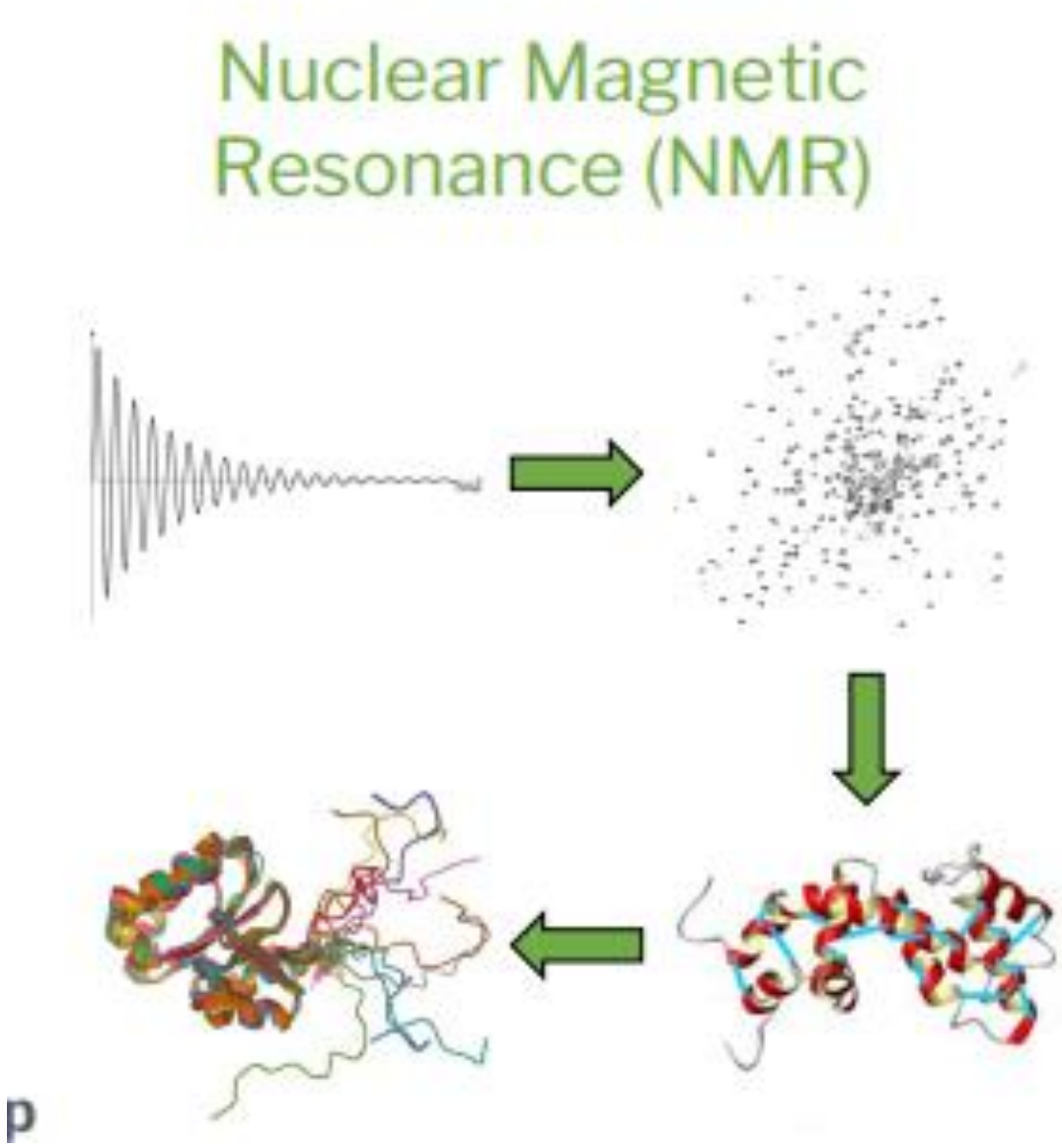
PDBID:1po5



Packing of glycolate oxidase (schematically)

https://bmrb.io/

NMR

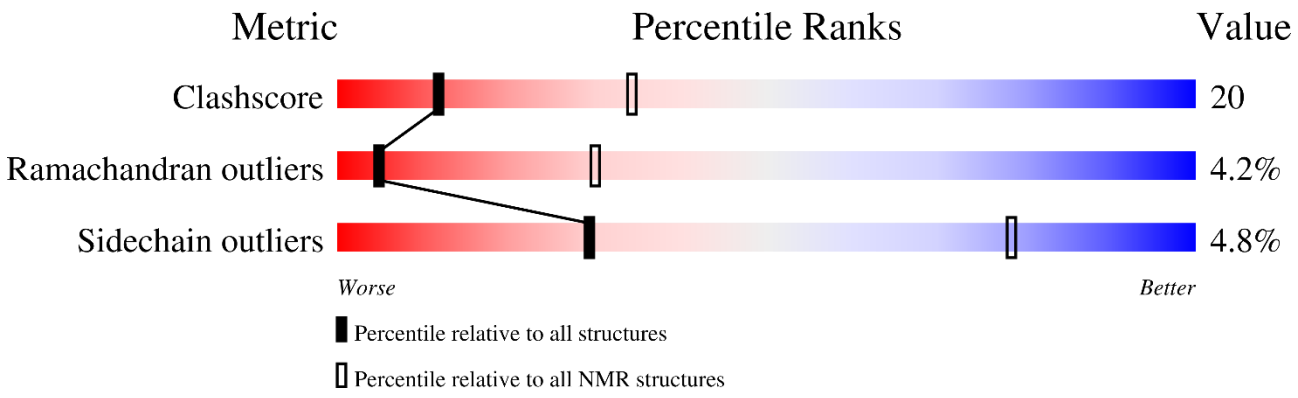


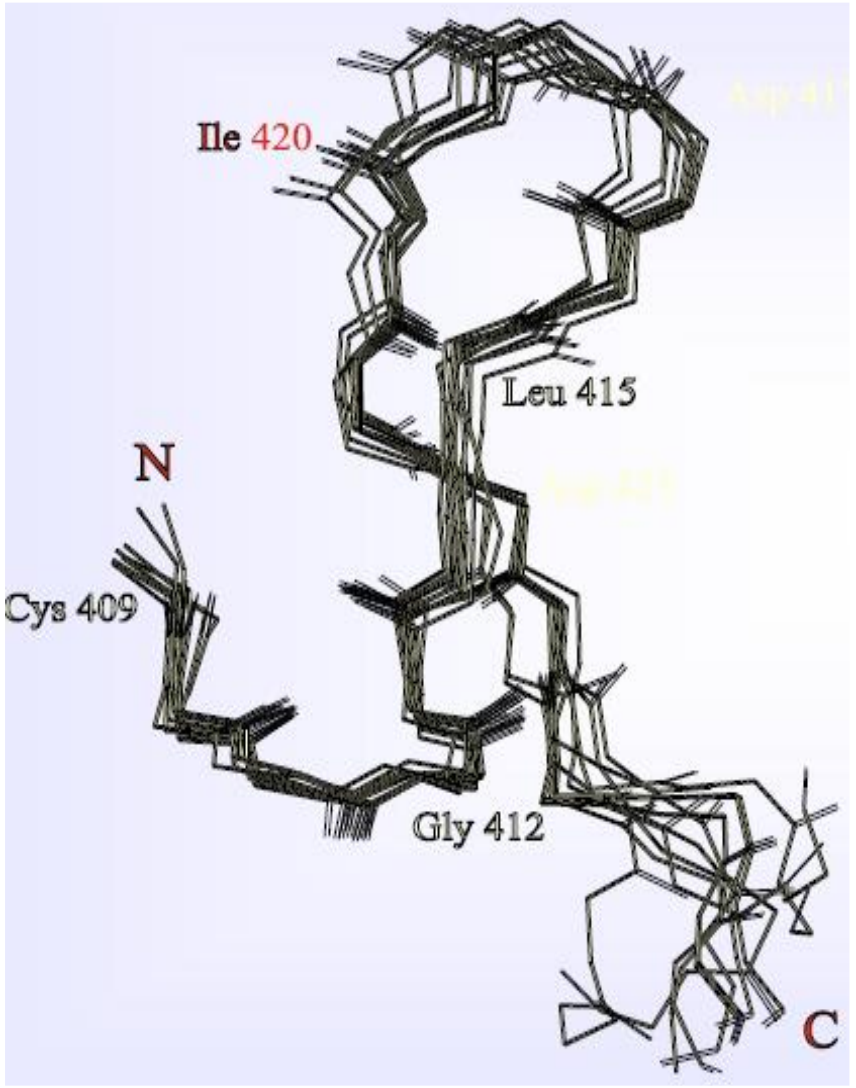
Structure Quality NMR

- NMR experiment

=> Multiple structures – e.g. 10 conformers

- Quality check wwPDB:
  - Stereochemistry – Ramachandran
  - Clashscore
  - Sidechain outliers



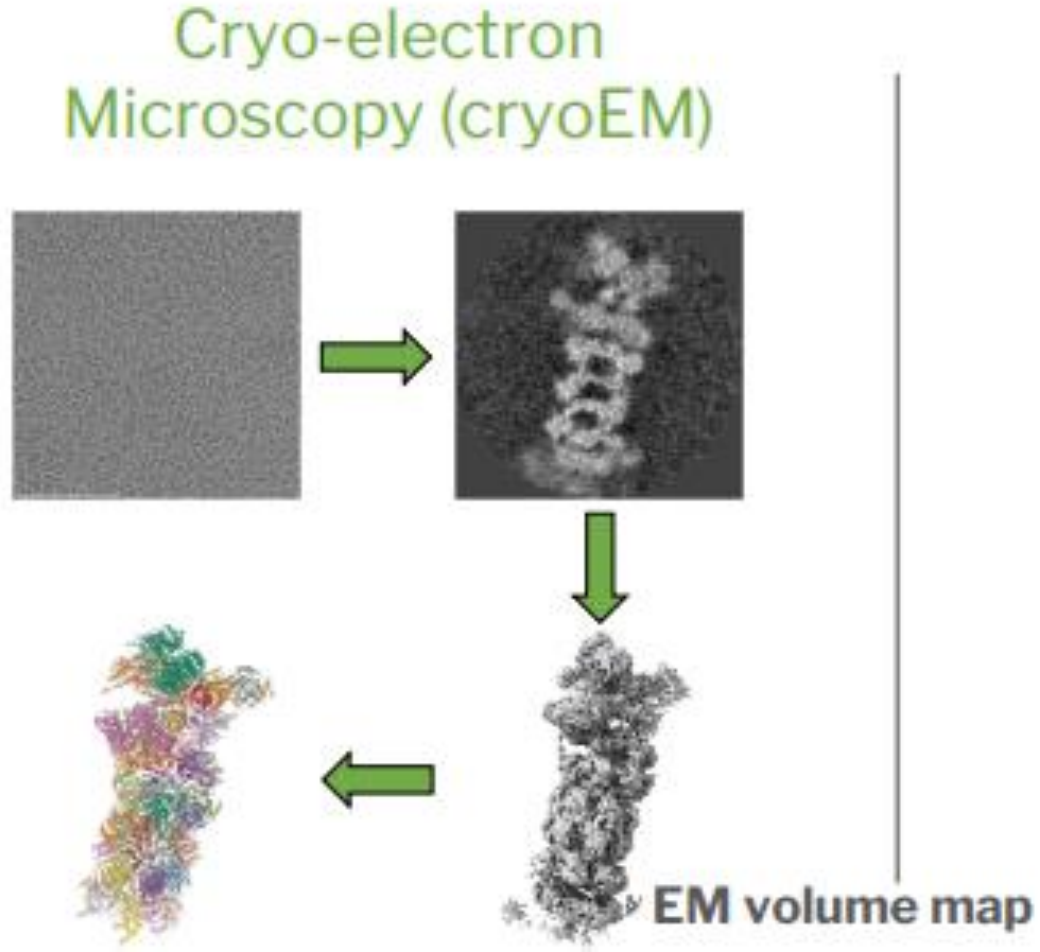


What we can get from NMR experiment

- Raw
  - Time-domain: 1D,2D,3D,4D spektrum
- Processed (FT)
  - frekvency domain
- NMR parametry
  - peak list
  - chemical shifts
  - 1H-1H NOE
  - J-couplings
  - residual dipolar couplings
  - NMR relaxation rates
- Derived data
  - NMR peak assignments
  - % expected in observed data
  - covalent structure
  - bond hybridizations
- Derived data
  - secondary structure
  - interatomic distances
  - torsion angles
  - hydrogen bonds
  - order parameters
  - solvent exposure
  - 3D structure
  - binding constant
  - pH titration parameters
  - hydrogen exchange rates
  - termodynamics and kinetics of structural rearangements
  - disordered regions

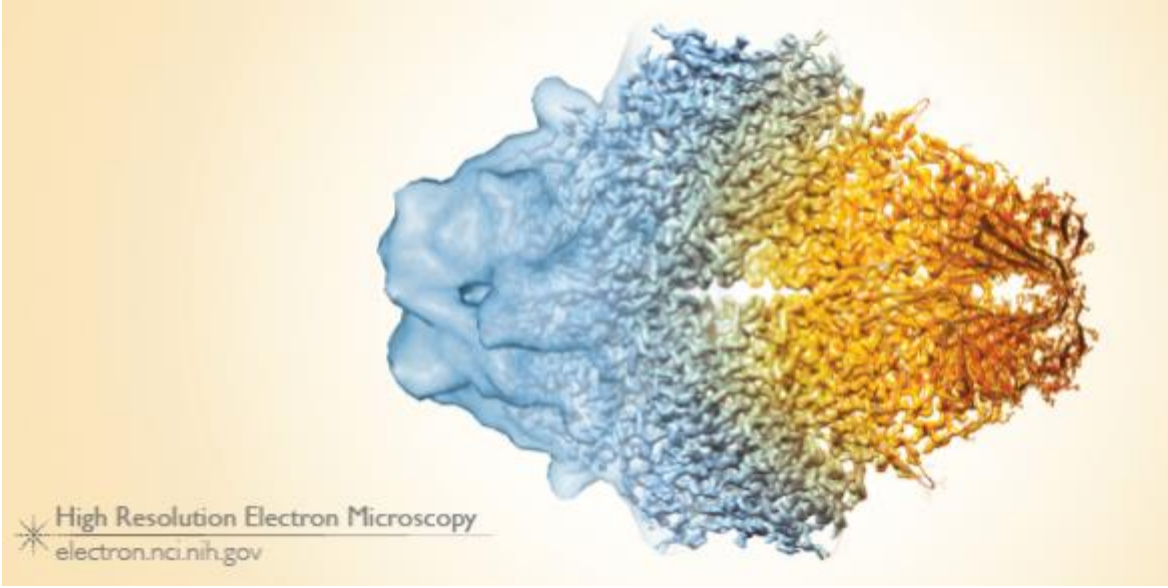
https://www.ebi.ac.uk/emdb/

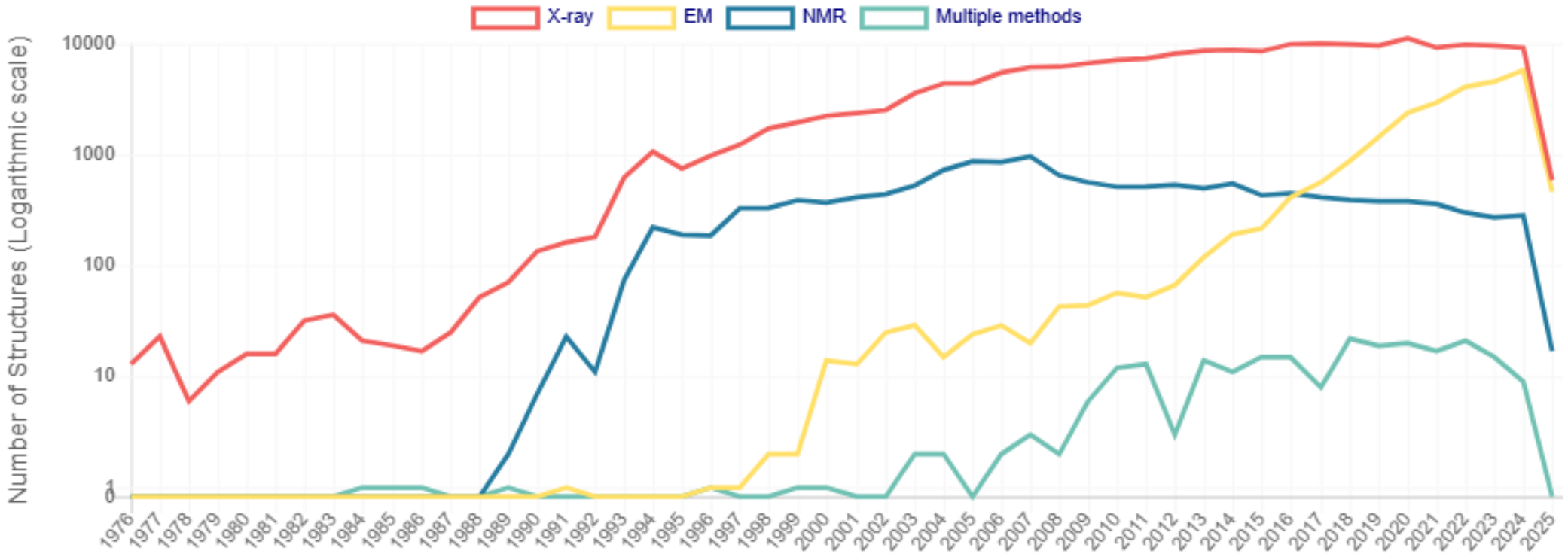
ELECTRON MICROSCOPY



CryoEM revolution

- Huge increase of resolution and amount of CryoEM data
- Similar quality measures to X-Ray





https://www.rcsb.org/stats/all-released-structures

EMDB

Protein Data Bank in Europe

Bringing Structure to Biology

EM resources

Share

EM Resources

- Home
- Statistics
- Validation
- EMDataBank
- EMPIAR
- Test data

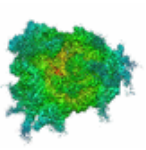
EMDB

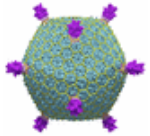
- Latest maps
- Latest headers
- Latest updates
- Search
- Browse
- FTP archive
- Deposit EM map/model
- EMDB data model

The Electron Microscopy Data Bank (EMDB) at PDBe

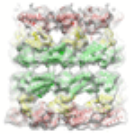
Quick access

Click on one of these categories:


Ribosome

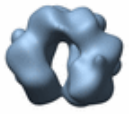

Virus


Phage

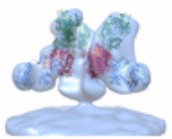

GroEL

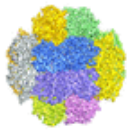

Microtubule


Polymerase

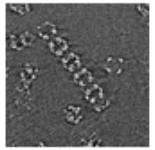

Helicase

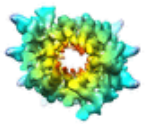

Human

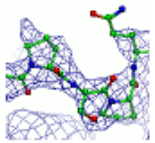

HIV


Entries with fitted models


Single particle


Tomography


Helical reconstruction


<5Å resolution

or enter 4-digit EMDB entry number: 1001 **Entry summary** **Visual analysis of map** **Volume viewer**

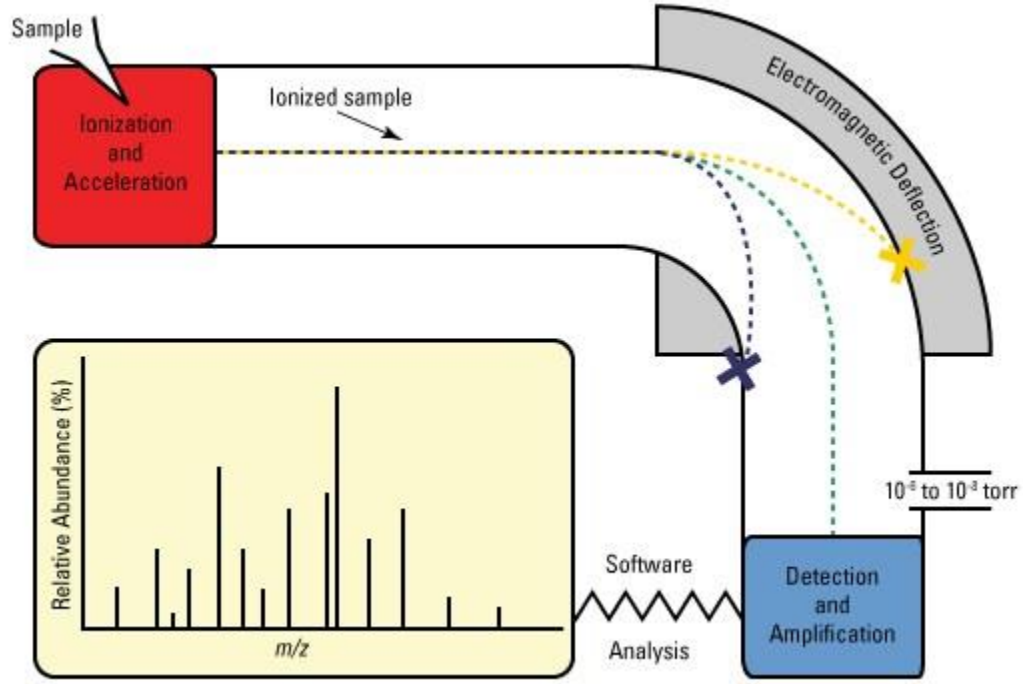
Introduction

The Electron Microscopy Data Bank (EMDB) is a public repository for electron microscopy density maps of macromo
and subcellular structures. It covers a variety of techniques, including single-particle analysis, electron tomogra
(2D) crystallography.
The EMDB was founded at EBI in 2002, under the leadership of Kim Henrick. Since 2007 it has been operated joi
and the Research Collaboratory for Structural Bioinformatics (RCSB PDB) as a part of EMDataBank which is fund
grant to PDBe, the RCSB and the National Center for Macromolecular Imaging (NCMI).

https://massbank.eu/ - small molecules

https://www.ebi.ac.uk/pride/ - proteins

MASS SPECTROMETRY



https://www.thermofisher.com/cz/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-mass-spectrometry.html

Mass spectrometry

- Proteomics
  - Determine protein structure, function, folding and interactions – **Crosslinks (X-MS)**
  - Identify a protein from the mass of its peptide fragments.
  - Detect specific post-translational modifications throughout complex biological mixtures using workflows for phosphoproteomics and protein glycosylation.
  - Quantitate proteins (relative or absolute) in a given sample.
  - Monitor enzyme reactions, chemical modifications and protein digestion.
- Drug Discovery
  - Determine structures of drugs and metabolites.
  - Screen for metabolites in biological systems.

Hu.map3

>25,000 mass spectrometry experiments

identify >15,000 human protein complexes

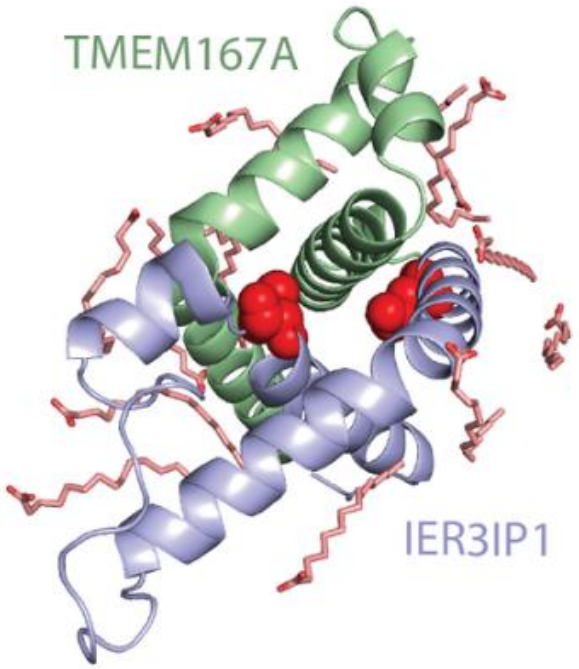
~75% of human proteins into physical contexts

protein co-variation data (ProteomeHD.2)

testable functional hypotheses for 472 uncharacterized proteins using AlphaFold modeling.

ebi.ac.uk/complexportal

humap3.proteincomplexes.org

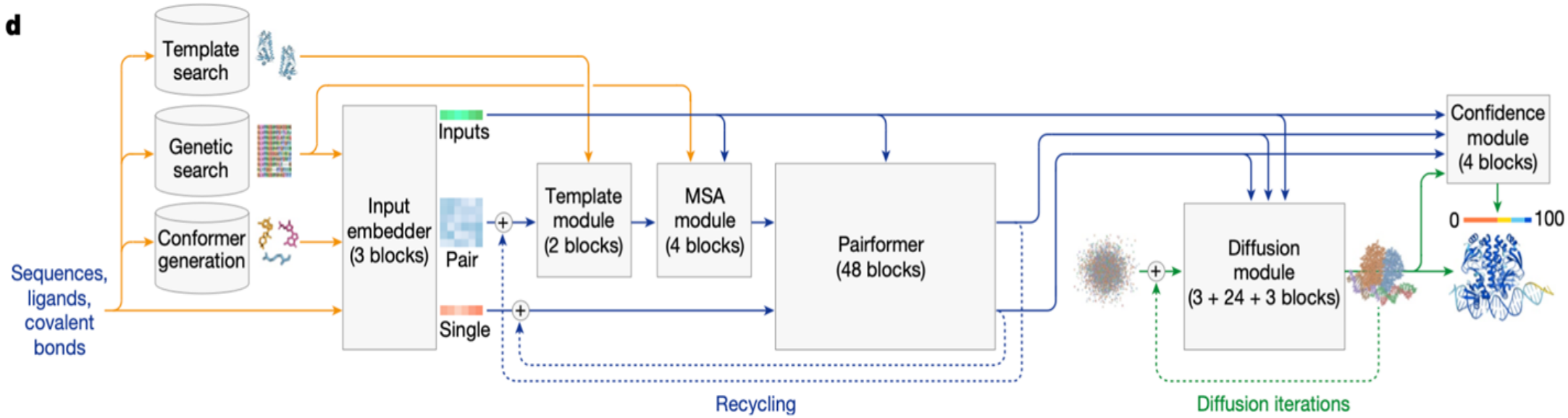


hu.MAP3.0 Score = 0.992

MED syndrome patient mutations V21G, L78P

ipTM = 0.77, pTM = 0.8

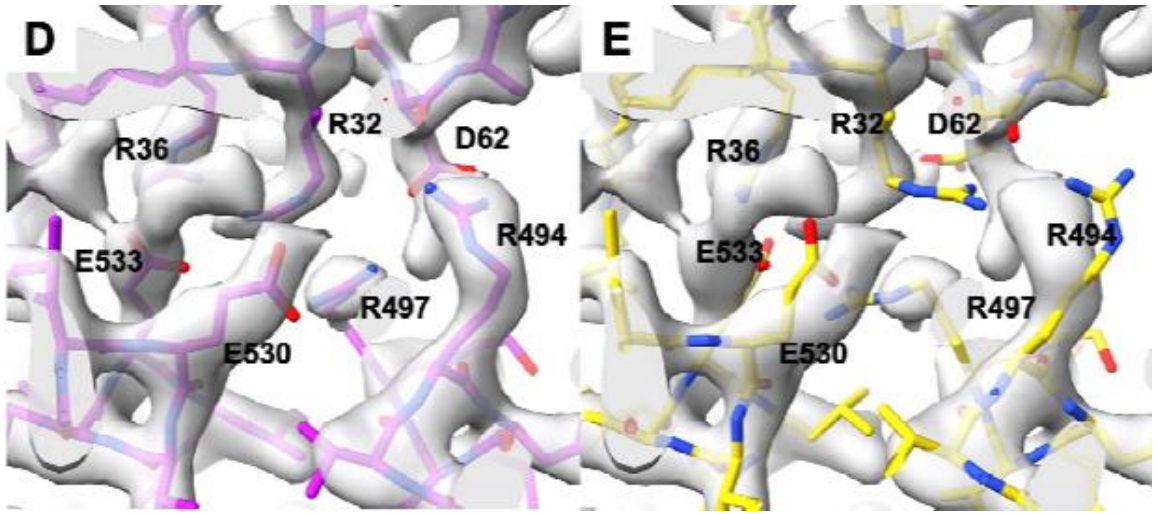
Fischer SN, et al. **hu.MAP3.0: Atlas of human protein complexes by integration of > 25,000 proteomic experiments** *BioRxiv* https://doi.org/10.1101/2024.10.11.617930



https://alphafoldserver.com/

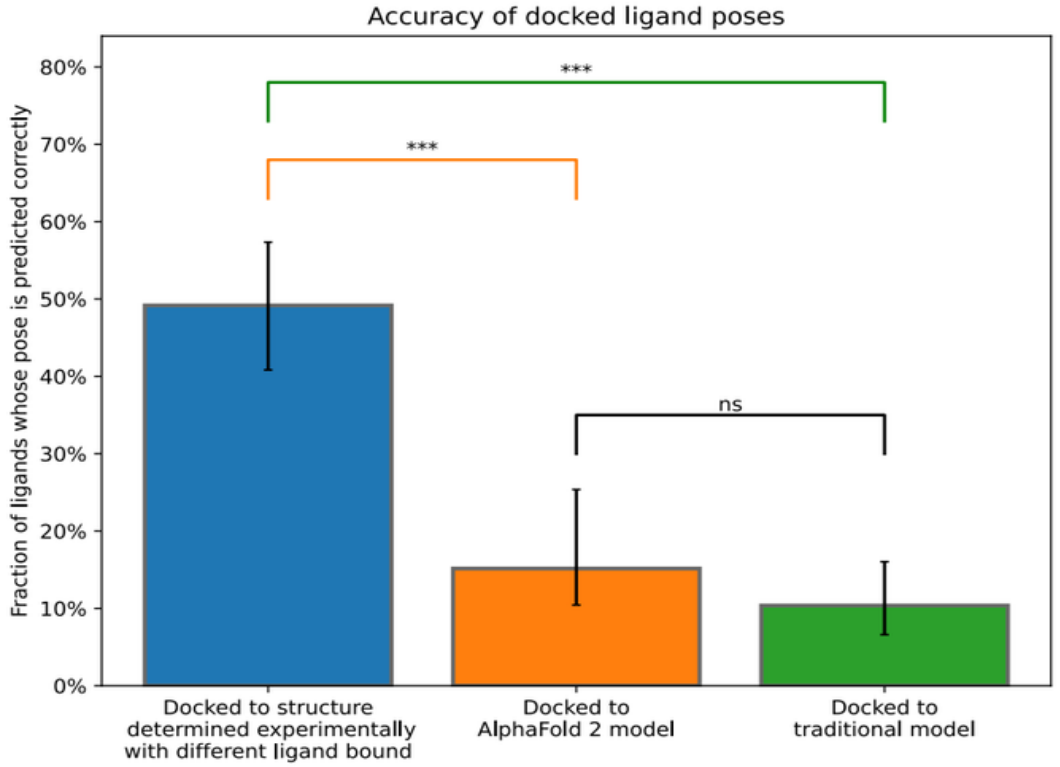
MODELLING ALPHAFOLD

How good are AlphaFold models for drug design?



AlphaFold predictions are valuable hypotheses, and accelerate but do not replace experimental structure determination

Thomas C. Terwilliger, Dorothee Liebschner, Tristan I. Croll, Christopher J. Williams, Airlie J. McCoy, Billy K. Poon, Pavel V. Afonine, Robert D. Oeffner, Jane S. Richardson, Randy J. Read, Paul D. Adams



How accurately can one predict drug binding modes using AlphaFold models?

Masha Karelina, Joseph J. Noh, Ron O. Dror

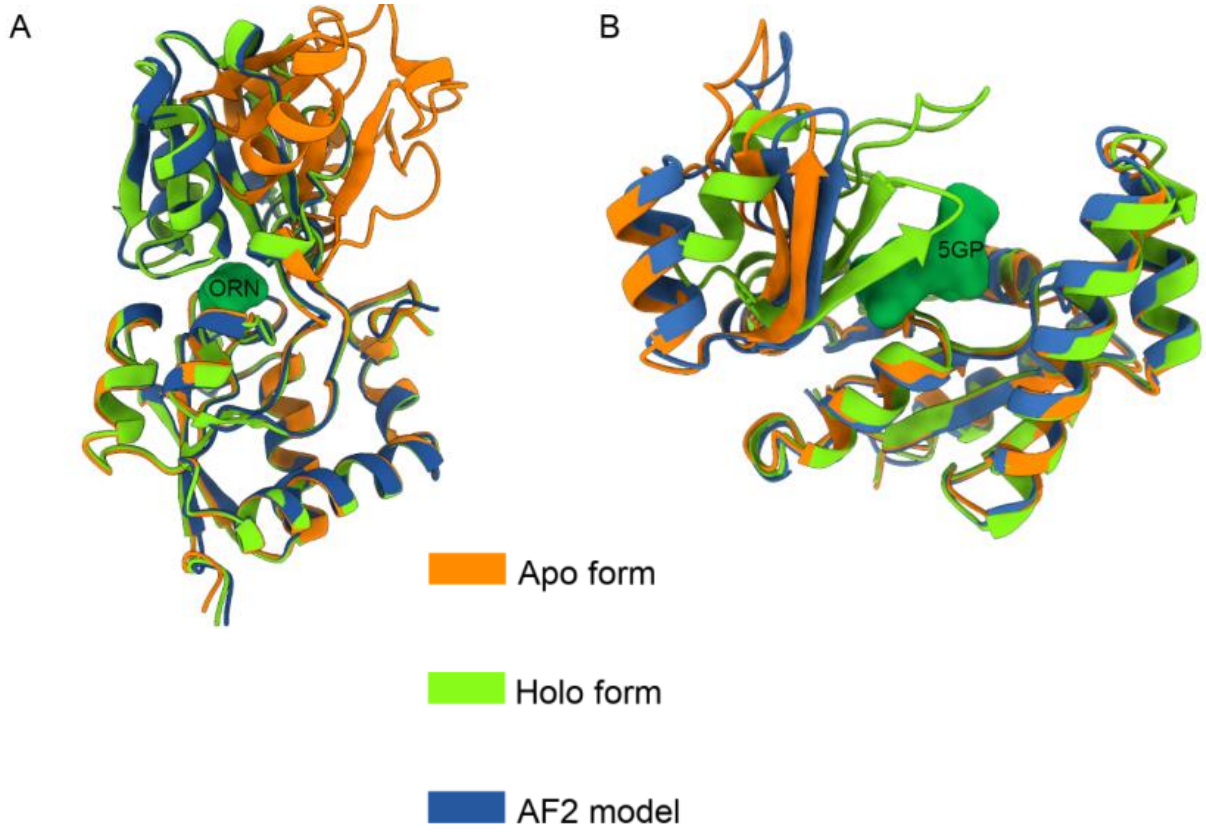
doi: https://doi.org/10.1101/2023.05.18.541346

This article is a preprint and has not been certified by peer review [what does this mean?].

AlphaFold models good enough for **drug design**?

- AlphaFold2 predicts **holo** protein in 70% => it can be used for drug designing
- pLDDT values in a single 3D model could be used to infer local conformational changes linked to ligand binding transitions.
- locally AlphaFold2 can be there - but it needs validation (as always)

- Good to combine with MD to optimize side-chain conformations



Impact of protein conformational diversity on AlphaFold predictions

Tadeo Saldaño, Nahuel Escobedo, Julia Marchetti, Diego Javier Zea, Juan Mac Donagh, Ana Julia Velez Rueda, Eduardo Gonik, Agustina García Melani, Julieta Novomisky Nechcoff, Martín N. Salas, Tomás Peters, Nicolás Demitroff, Sebastian Fernandez Alberti, Nicolas Palopoli, Maria Silvina Fornasari, Gustavo Parisi

doi: https://doi.org/10.1101/2021.10.27.466189

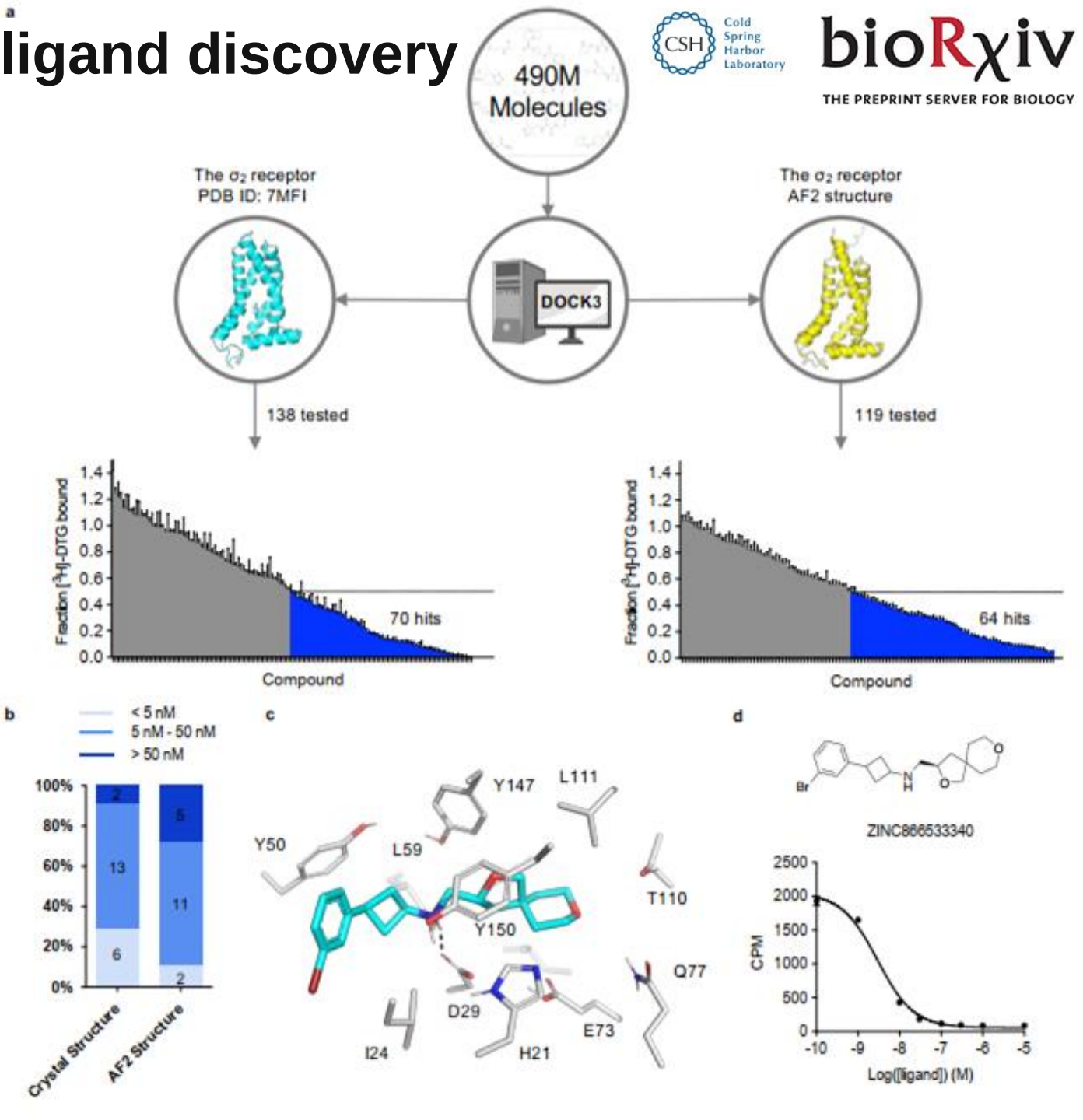
AlphaFold2 structures template ligand discovery

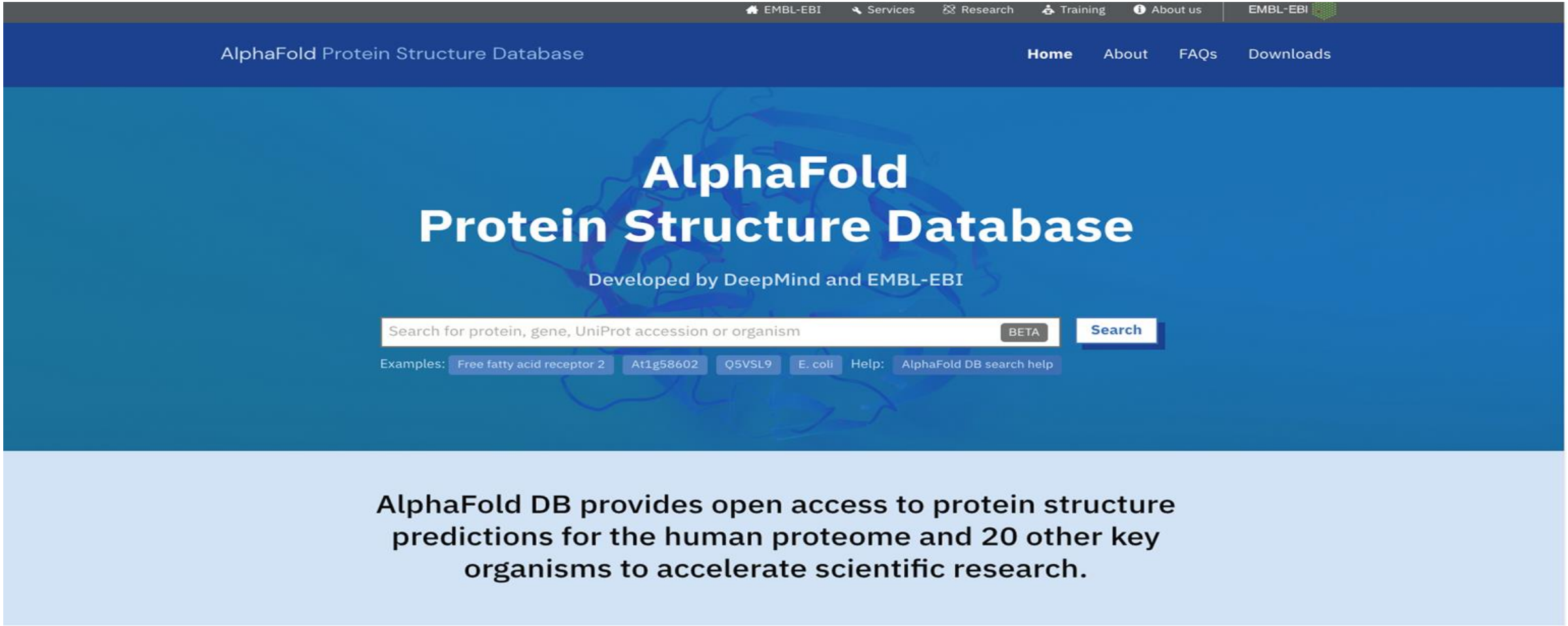
Retrospective docking screens against the σ_2 and the 5-HT2A receptors, the AF2 structures **struggled**

Prospective docking against the AF2 models => **similar hit rates** for both receptors versus docking against experimentally-derived structures

Jiankun Lyu, Nicholas Kapolka, Ryan Gumpper, Assaf Alon, Liang Wang, Manish K. Jain, Ximena Barros-Álvarez, Kensuke Sakamoto, Yoojoong Kim, Jeffrey DiBerto, Kuglae Kim, Tia A. Tummino, Sijie Huang, John J. Irwin, Olga O. Tarkhanova, Yurii Moroz, Georgios Skiniotis, Andrew C. Kruse, Brian K. Shoichet, Bryan L. Roth

doi: https://doi.org/10.1101/2023.12.20.572662





"This will be one of the most important datasets since the mapping of the Human Genome."
Professor Ewan Birney
EMBL Deputy Director General and EMBL-EBI Director

https://www.alphafold.ebi.ac.uk/

Complete structures of 48 model organism proteomes

AlphaFold DB currently provides predicted structures for the 48 organisms (including human), as well as the majority of Swiss-Prot. **>200 M structures**

Compressed prediction files for model organism proteomes:

Species	Common Name	Reference Proteome	Predicted Structures	Download
Arabidopsis thaliana	*Arabidopsis*	UP000006548	27,434	Download (3,678 MB)
Caenorhabditis elegans	Nematode worm	UP000001940	19,694	Download (2,626 MB)
Candida albicans	*C. albicans*	UP000000559	5,974	Download (974 MB)
Danio rerio	Zebrafish	UP000000437	24,664	Download (4,180 MB)

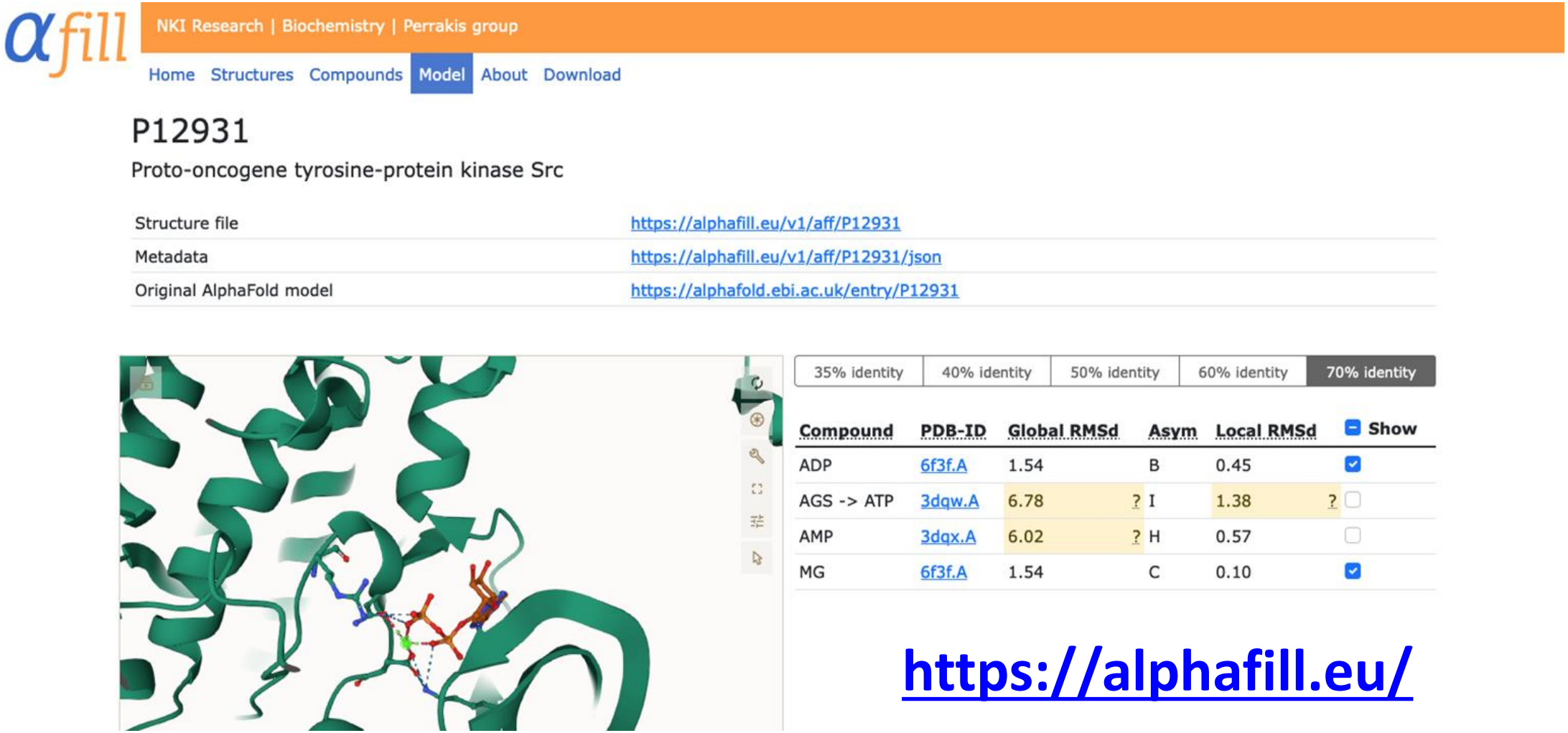
Compressed prediction files for global health proteomes:

Species	Common Name	Reference Proteome	Predicted Structures	Download
Ajellomyces capsulatus	Ajellomyces capsulatus	UP000001631	9,199	Download (1,351 MB)
Brugia malayi	Brugia malayi	UP000006672	8,743	Download (1,274 MB)
Campylobacter jejuni	*C. jejuni*	UP000000799	1,620	Download (173 MB)
Cladophialophora carrionii	Cladophialophora carrionii	UP000094526	11,170	Download (1,716 MB)

Compressed prediction files for Swiss-Prot:

File type	Predicted Structures	Download
Swiss-Prot (CIF files)	542,380	Download (36,896 MB)
Swiss-Prot (PDB files)	542,380	Download (26,935 MB)

Alphafold can be **Alphafill**-ed with **ligands + cofactors**



https://alphafill.eu/

Hekkelman, M.L., de Vries, I., Joosten, R.P. *et al.* AlphaFill: enriching AlphaFold models with ligands and cofactors. *Nat Methods* 20, 205–213 (2023). https://doi.org/10.1038/s41592-022-01685-y

Alphafold tells you where is it right!

SNW domain-containing protein 1

AlphaFold structure prediction

Download | PDB file | mmCIF file | Predicted aligned error

pLDDT – local confidence
PAE – global confidence
pTM – predicted template model
ipTM – interface predicted TM

Information

Protein	SNW domain-containing protein 1
Gene	SNW1
Source organism	Homo sapiens go to search
UniProt	Q13573 go to UniProt
Experimental structures	17 structures in PDB for Q13573 go to PDBe-KB
Biological function	(Microbial infection) Proposed to be involved in transcriptional activation by EBV EBNA2 of CBF-1/RBPJ-repressed promoters. go to UniProt

3D viewer

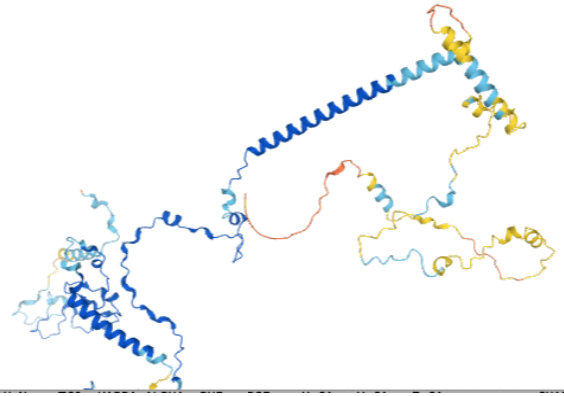
Model Confidence:

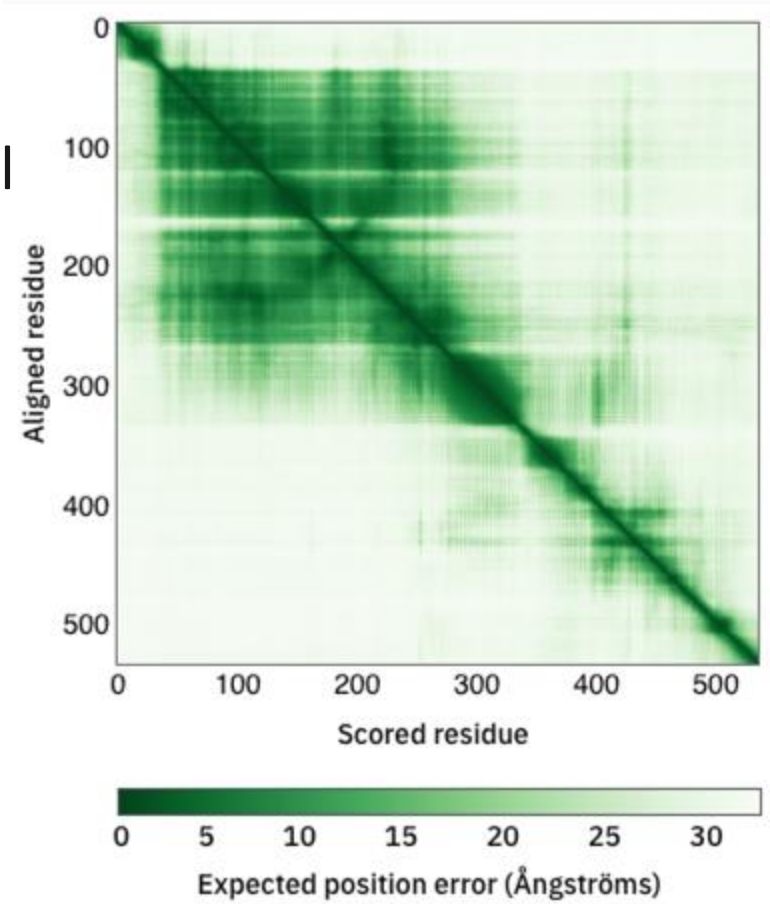
- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

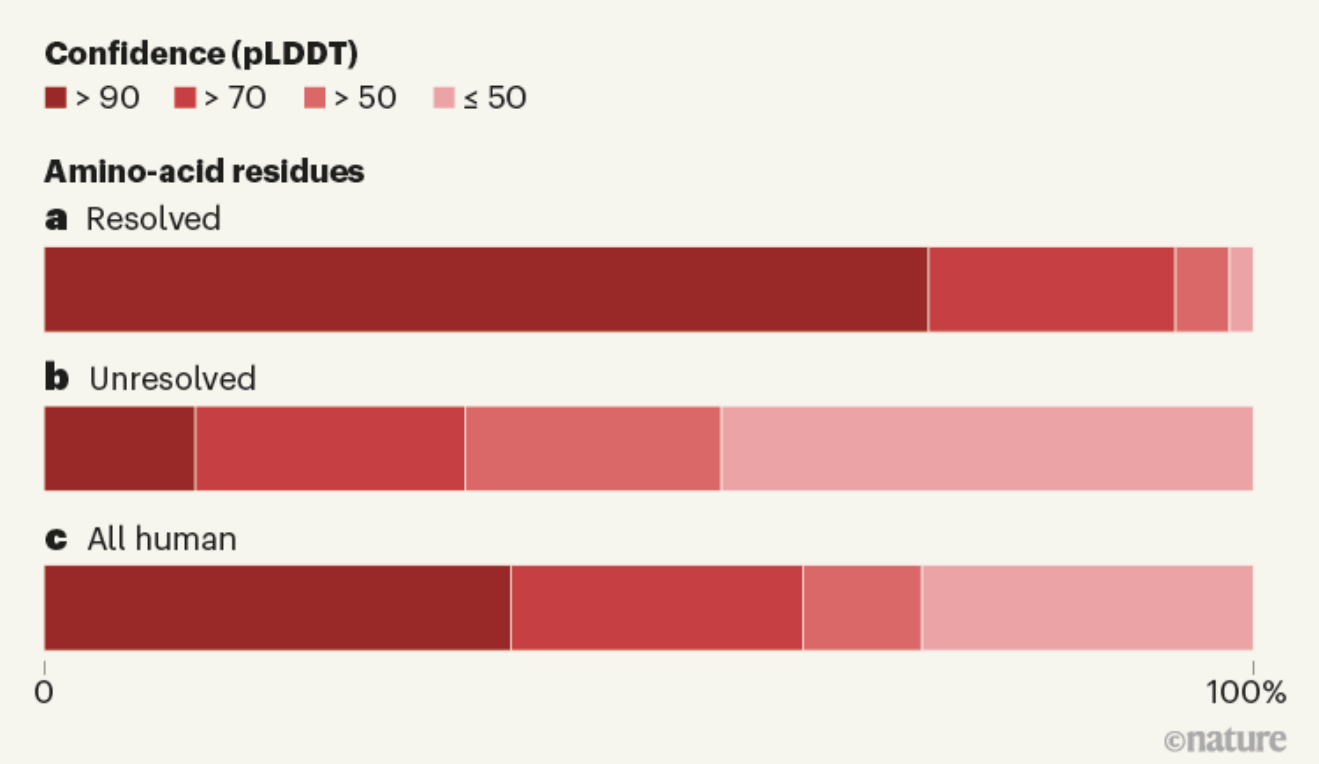
Sequence of AF-Q13573-... 1: SNW do... A

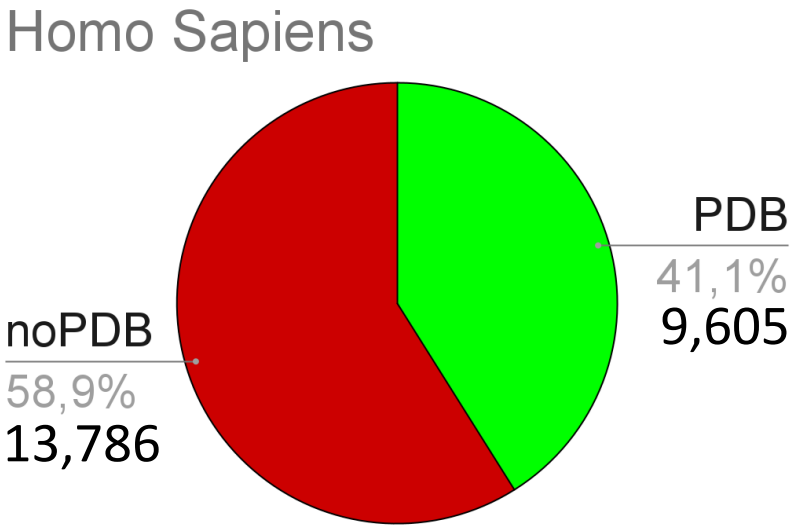
```
MALTSFLPAPTQLSQDQLEAEEKARSQRSRQTSLVSSRREPPPYGYRKGWIPRLLEDFGDGGAFPEIHVAQYPLDMGRKKKMSNALAIQVDSEGKIKYDAIARQGQSKDKVIYSKYTDLVPKEV
MNADDPDLQRPDEEAIKEITEKTRVALEKSVSQKVAAAMPVRAADKLAPAQYIRYTPSQQGVAFNSGAKQRVIRMVEMQKDPMEPPRFKINKKIPRGPPSPPAPVMHSPSRKMTVKEQQEWKIP
PCISNWKNAKGYTIPLDKRLAADGRGLQTVHINENFAKLAEALYIADRKAREAVEMRAQVERKMAQKEKEKHEEKLREMAQKARERRAGIKTHVEKEDGEARERDEIRHDRRKERQHDRNLSRA
```





How good are the predictions of human proteins?





pLDDT

- quality metrics
- measure of disorder

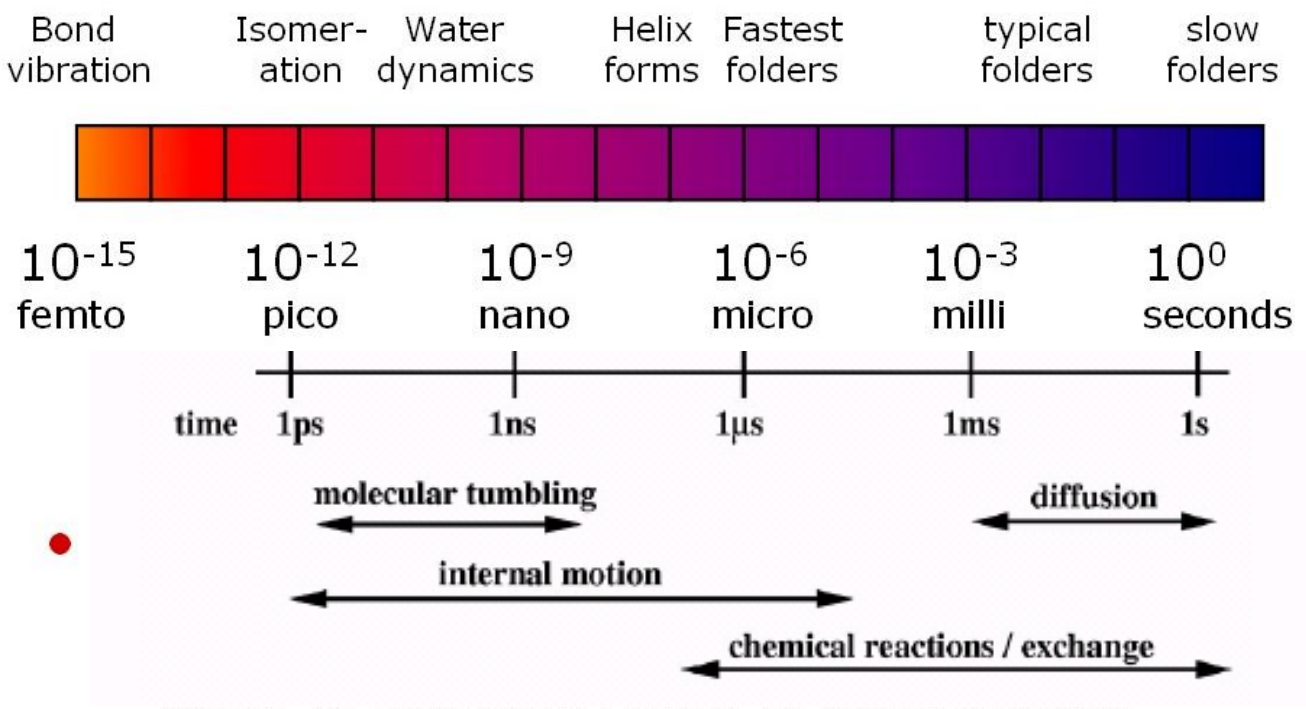
pLDDT - per-residue estimate of its confidence on a scale from 0 - 100
model's predicted score on the lDDT-Cα metric (local superposition-free score for comparing protein structures and models using distance difference tests).

Tunyasuvunakool, K., Adler, J., Wu, Z. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* 596, 590–596 (2021). https://doi.org/10.1038/s41586-021-03828-1

Unstructured part of proteins

DISORDER

Time Scale of Protein Movement



- **16 order of magnitude range**
  - Femtosecond timesteps
  - Need to simulate micro to milliseconds

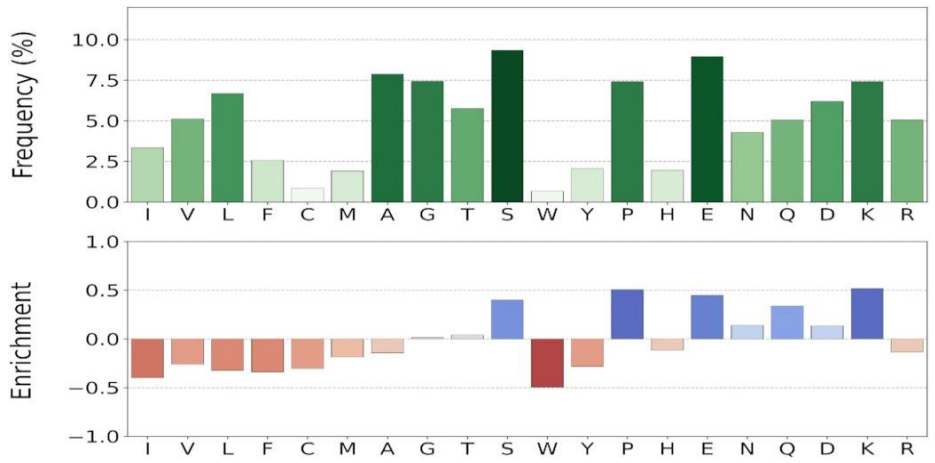
http://pande.stanford.edu

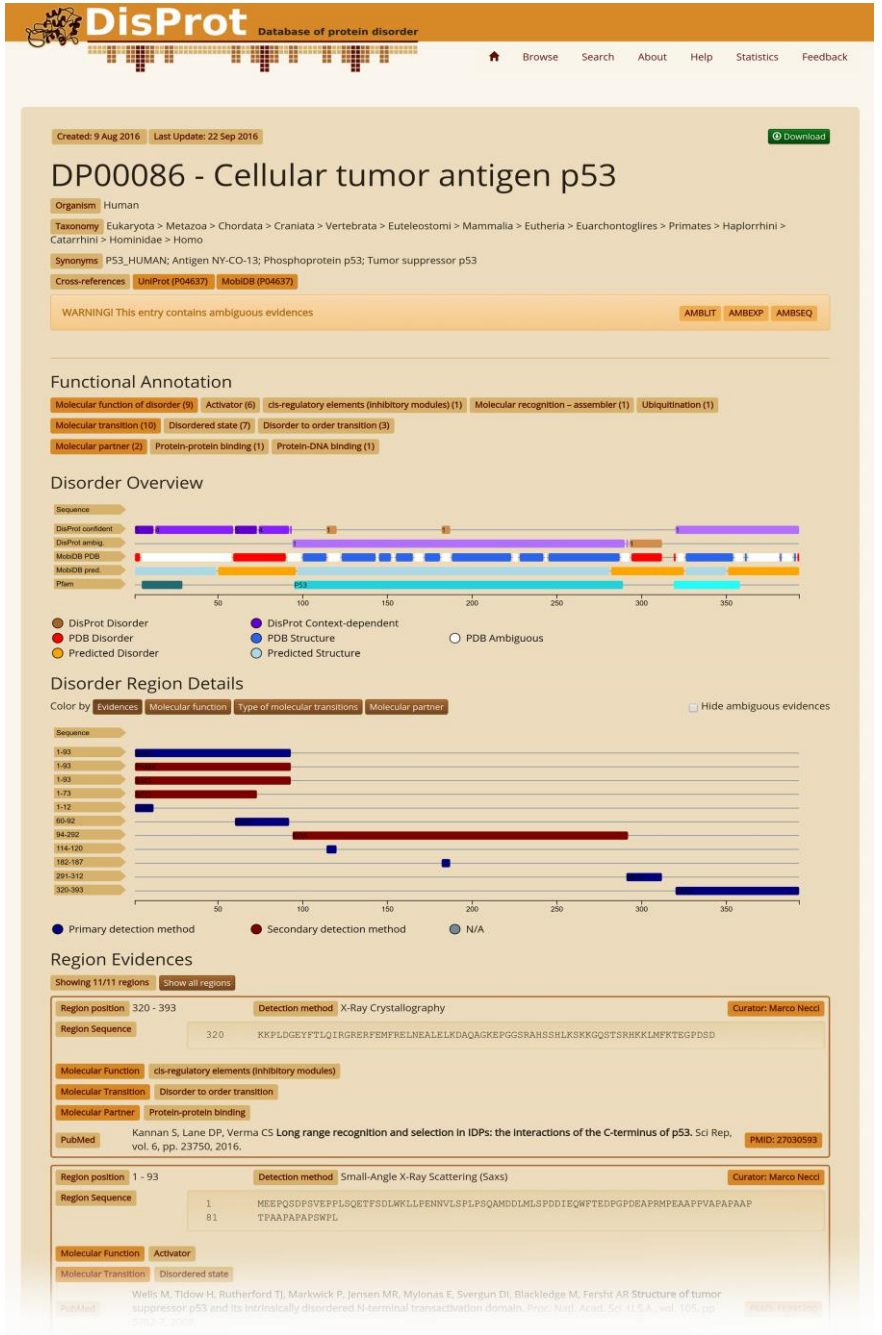
Disorder

- Invisible – no structure
- - temperature B-factor v Xray, NMR ensemble, DISOPRED predictions
- Quite conserved arrangement within a protein family
- **Intristically Disordered Proteins - IDP**
  - Function
    - molecular recognition (promiskuitní)
    - molecular assembly (viral capsids)
    - protein modification
    - entropic chain activities

Disprot - Database of disorder in proteins

- Amino acid composition of DisProt disordered regions.
- sorted by the Kyte-Doolittle hydrophobicity scale.
- enrichment is calculated and normalized over the TrEMBL database frequencies (2021_03).





- *Nucleic Acids Res*, Volume 50, Issue D1, 7 January 2022, Pages D480–D487, https://doi.org/10.1093/nar/gkab1082
- *Nucleic Acids Res*, Volume 45, Issue D1, January 2017, Pages D219–D227, https://doi.org/10.1093/nar/gkw1056

OXFORD UNIVERSITY PRESS

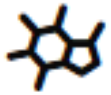
Aggregated views of proteins and ligands

https://www.ebi.ac.uk/pdbe/pdbe-kb

PDBe-KB
Protein Data Bank in Europe Knowledge Base

- **Aggregated views of proteins**

Structures



Small-molecules

Macromolecular Interactions

Functional Annotations

- API
  - https://www.ebi.ac.uk/pdbe/graph-api/pdbe_doc/
- Component library
  - https://github.com/PDBe-KB?q=component

Mediator of DNA damage checkpoint protein 1

Gene: MDC1

Organism: *Homo sapiens (Human)*

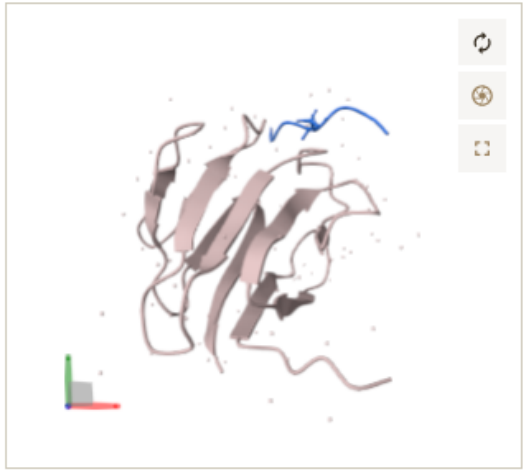
Synonyms: KIAA0170, NFBD1

Uniprot: Q14676 go to UniProt

Biological function: Histone reader protein required for checkpoint-mediated cell cycle arrest in response to DNA damage within both the S phase and G2/M phases of the cell cycle (PubMed:12475977, PubMed:12499369, PubMed:12551934, PubMed:12607003, PubMed:12607004, PubMed:12607005, PubMed:12611903, PubMed:14695167, **[show more]** go to UniProt

Representative structures for UniProt Q14676

PDB chains with highest data quality, coverage and best resolution



PDB chain shown: 3unn B go to PDBe

UniProt residues 1 - 8

Coverage: unavailable

No superposed structures for this region

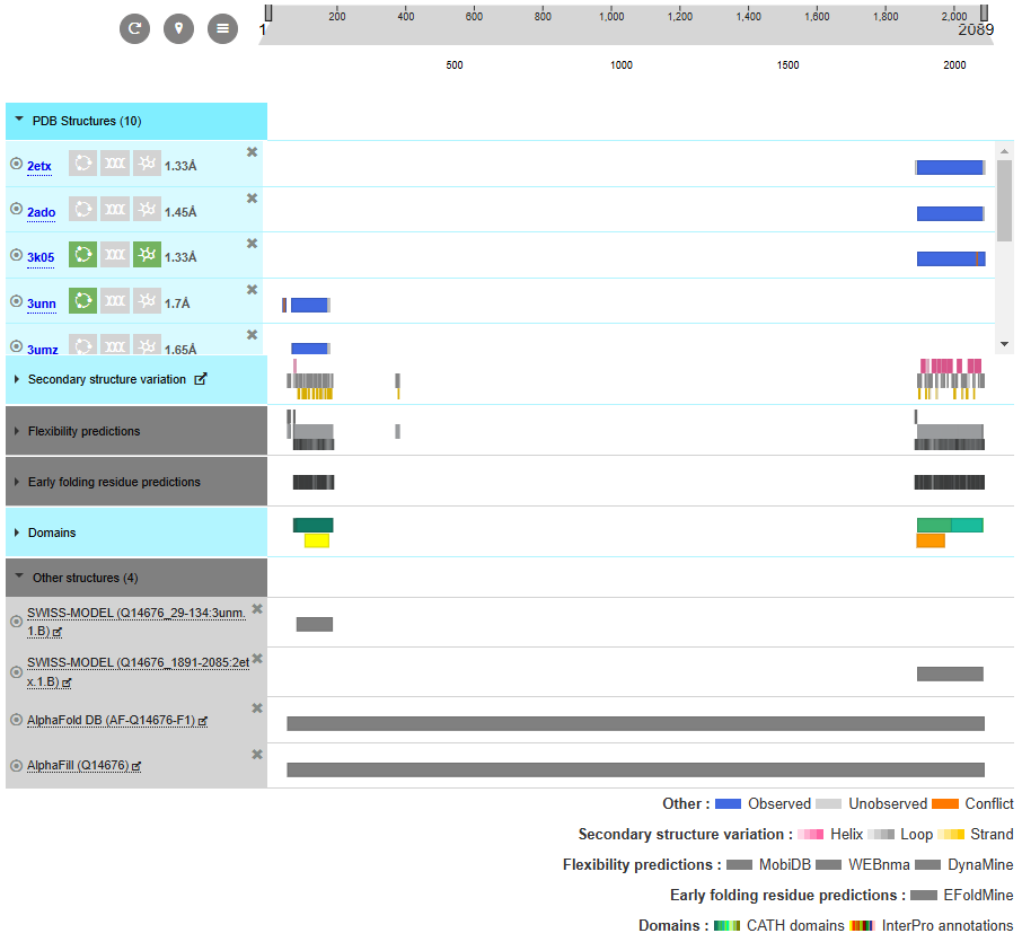
Click on the icons below to view the relevant page:

10 Structures	2 Ligands	4 Interactions	Annotations	0 Similarity	345 Publications
Download	Download	Download			
3D view of superposed structures	3D view of superposed ligands				

PDBe-KB

Protein Data Bank in Europe Knowledge Base

ProtVista



HEM

Cofactor-like | Drug-like | Reactant-like

PROTOPORPHYRIN IX CONTAINING FE

PDBe-KB Ligands
Protein Data Bank in Europe - Knowledge Base

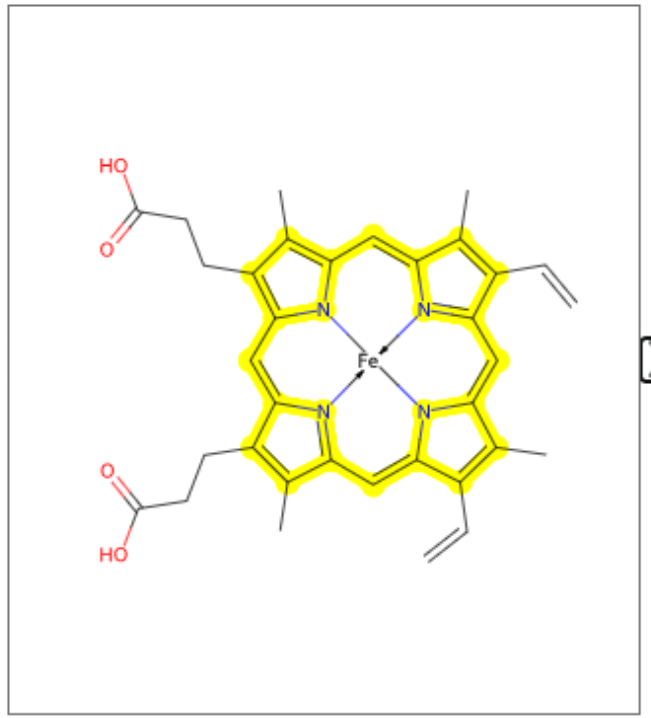
View 3D | Download as

Description

Synonyms	HEME, Heme, HEME B, PROTOHEME, HEME IRON, PROTOHEME IX
Formula	C34 H32 Fe N4 O4
IUPAC InChI	InChI=1S/C34H34N4O4.Fe/c1-7-21-17(3)25-13-26-19(5)23(9-11-33(39)40)31(37-26)16-32-24(10-12-34(41)42)20(6)28(38-32)15-30-... Show more
IUPAC InChIKey	KABFMIBPWCXCRK-RGGAHWMASA-L
SMILES	Cc1c2n3c(c1CCC(=O)O)C=C4C(=C(C5=[N]4[Fe]36[N]7=C(C=C8N6C(=C5... Show more
	Source OpenEye
First observed in	3ia3

View Atoms | View Bonds

Overall view, and highlighted scaffolds and fragments



Displayed: 3 / 5

Porphin-like fragment highlighted in yellow

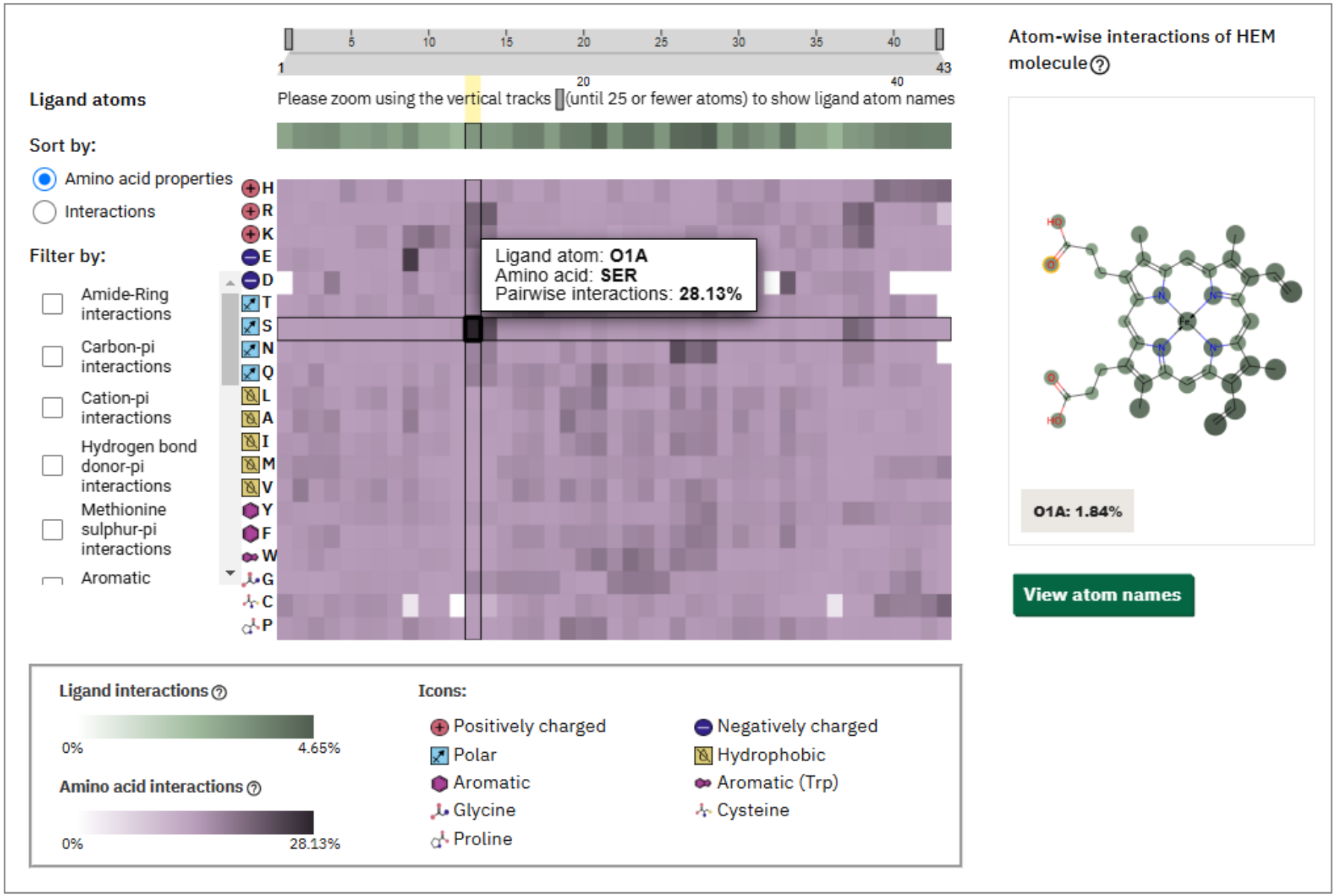
SMILES INCHIKey

https://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/HEM

Interaction statistics

Interaction statistics shows the summary of aggregated protein-ligand interaction data of HEM from 14370 ligand instances in 1049 PDB structures and 7011 PDB chains. The protein-ligand interactions are computed using PDBe Arpeggio

Download all interactions Documentation



Take home message

- Each method has its advantages and disadvantages
- X-ray
  - crystal (contacts), inner electrons – electron map
  - **R <2.5Å** is ok for drug design; worse resolution -> electron envelope
  - size is not restricted, phase problem, static
- NMR - size restriction, dynamical information possible – verification of models
- EM - electron envelopes (maps) > usually worse atomic resolution inside
  - good for protein complexes
- MS + FRET - distances only, need model
- AlphaFold
  - Easy to run, caution on the quality, drug design itself complicated
  - Visualize disorder
- PDBe-KB – aggregated view on protein and ligands

THANK YOU FOR YOUR ATTENTION

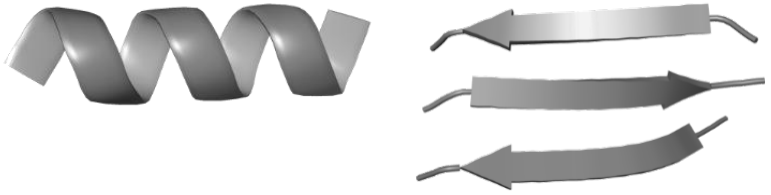
Questions?

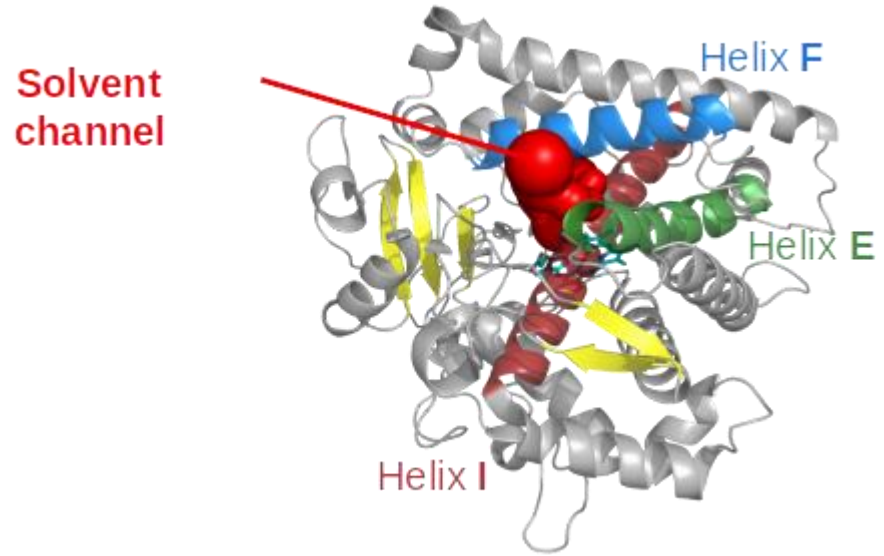
UNUSED SLIDES

STRUCTURE ANALYSIS

Secondary structure

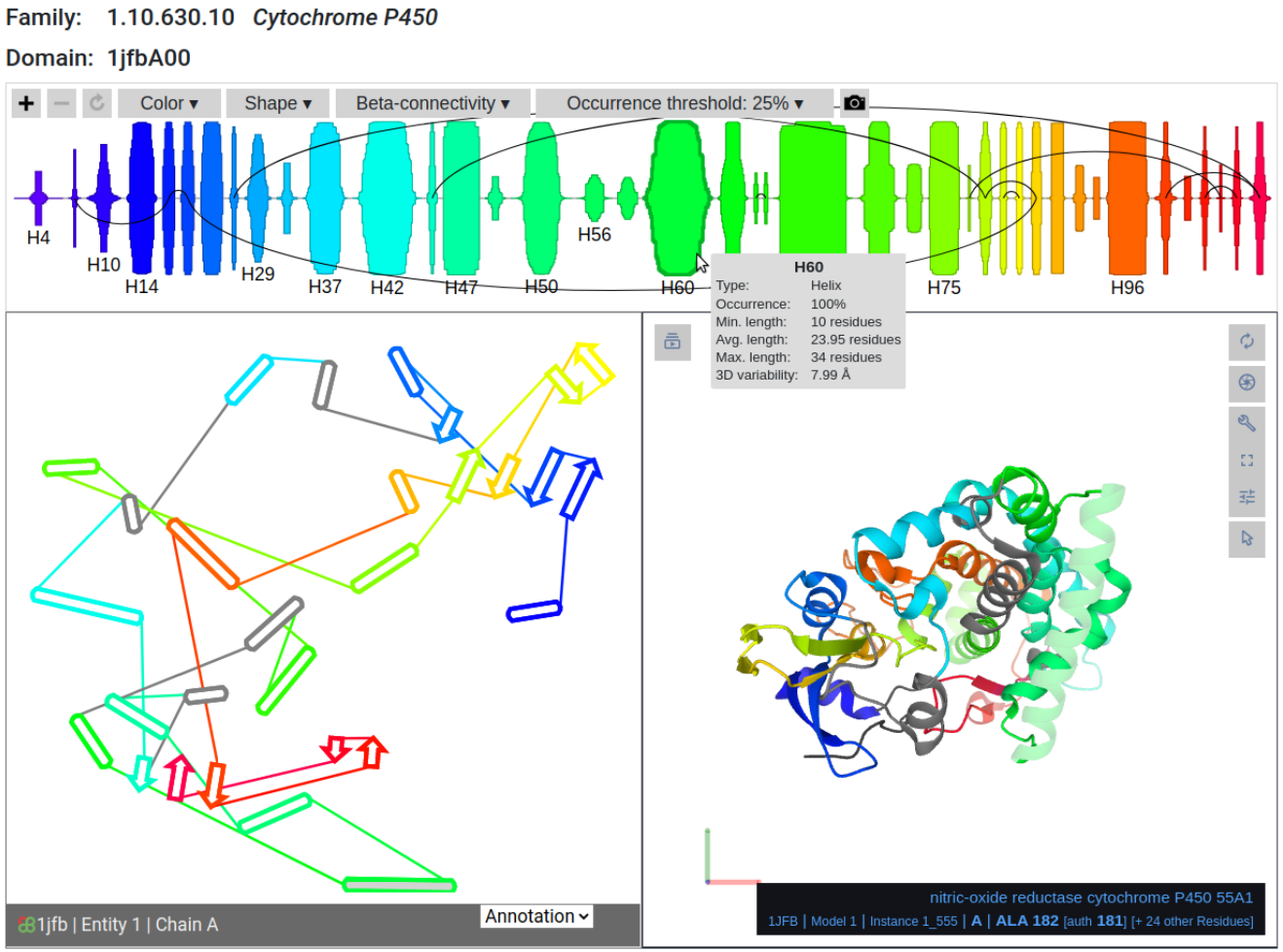
Helices and β-strands = Secondary Structure Elements (SSEs)





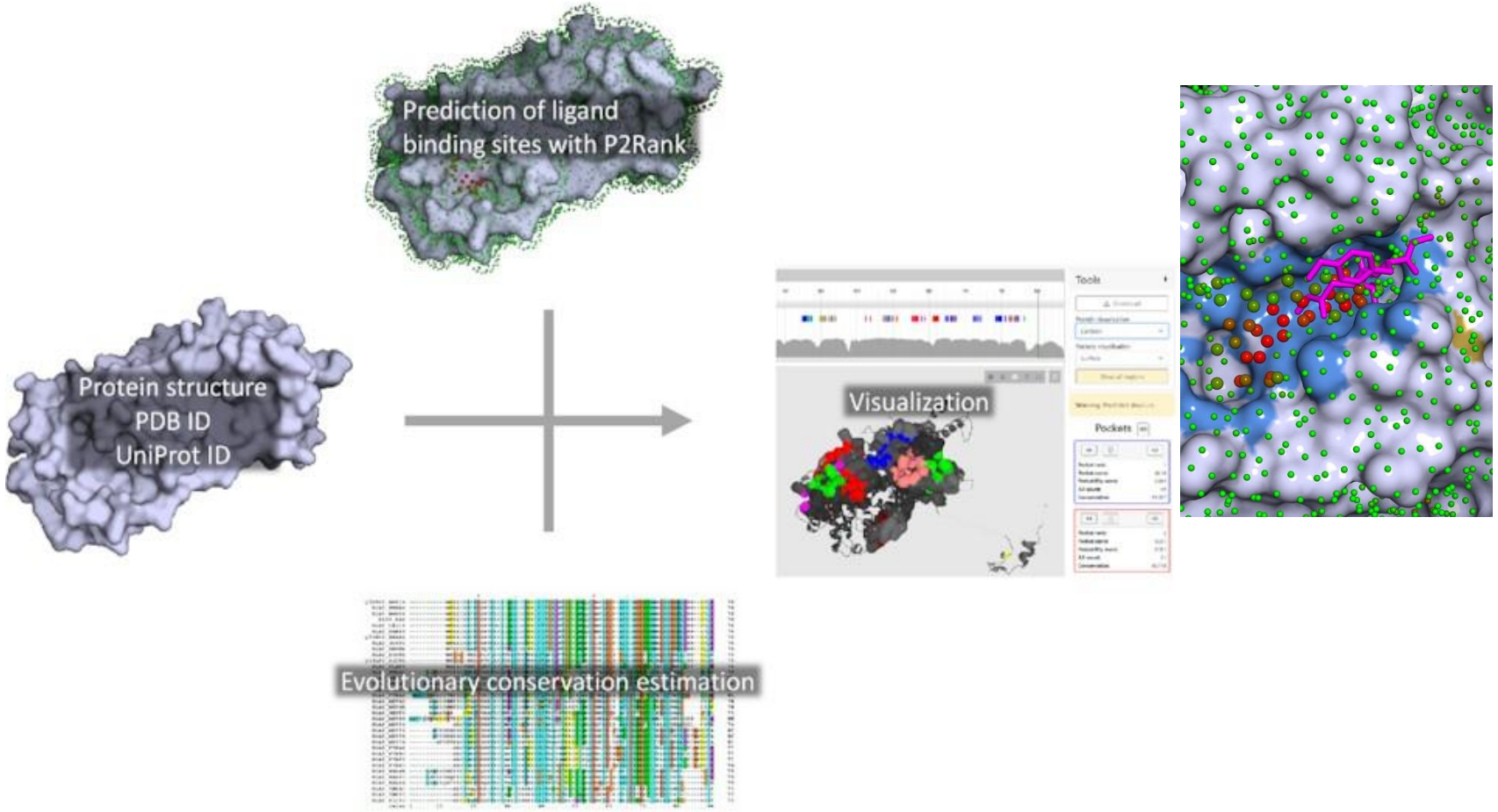
OverProt Server – Interactive view

- 1D of the family linked to 2D and 3D of a domain



Midlik A, Hutařová Vařeková I, Hutař J, Chareshneu A, Berka K, Svobodová R: **OverProt**: secondary structure consensus for protein families, *Bioinformatics*, 38(14), July 2022, 3648–3650

Binding Site prediction: P2Rank - PrankWeb



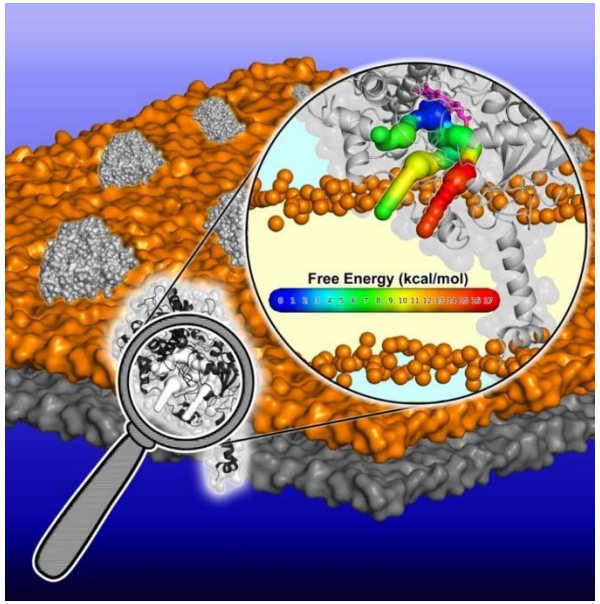
Each point represents its local chemical neighborhood –
- predicted ligandability score (0 = green to 1 = red)
- clustered to form predicted binding sites

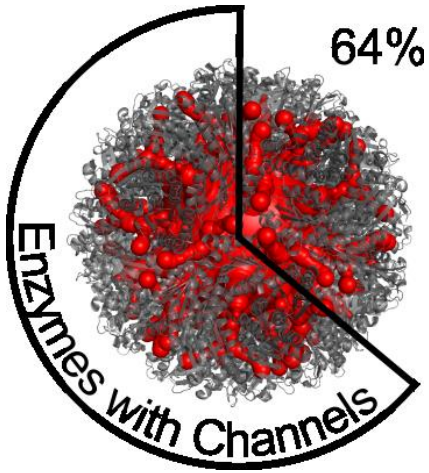
- https://prankweb.cz/
- https://github.com/cusbg/p2rank-framework

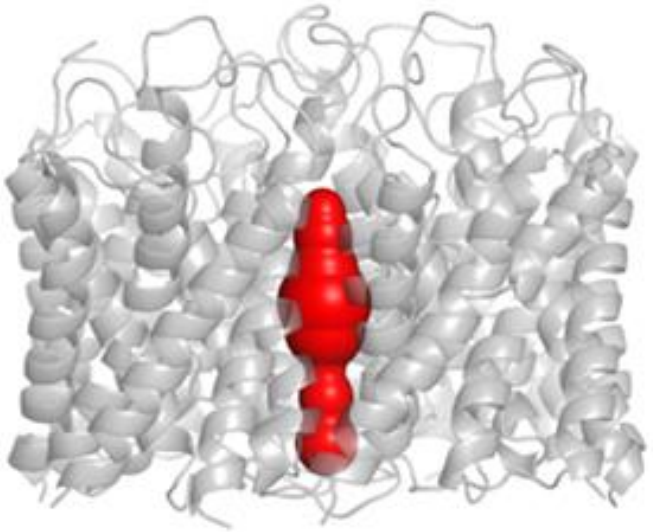
Nucleic Acids Res, Volume 50, Issue W1, 5 July 2022, Pages W593–W597, https://doi.org/10.1093/nar/gkac389

Porous Pathways in Proteins

- Channels/Tunnels
  - Connect active site with exterior
- Pores
  - Running through the structure







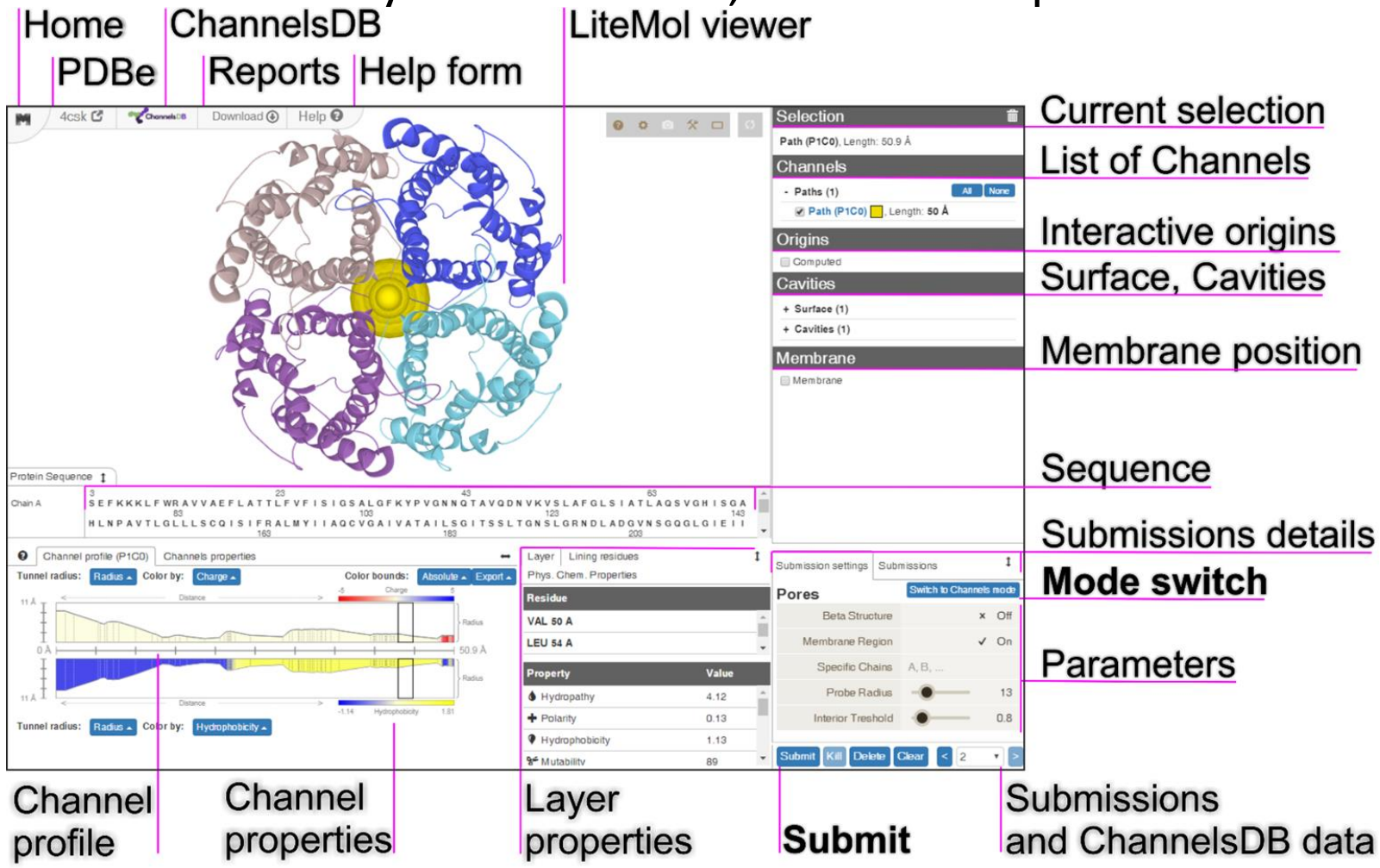
- Pathways have important function of gate-keeping

- Paloncýová M, Navrátilová V, Berka K, Laio A, Otyepka M: Role of Enzyme Flexibility in Ligand Access and Egress to Active Site – Bias-Exchange Metadynamics Study of 1,3,7-Trimethyluric Acid in Cytochrome P450 3A4 *J. Chem. Theory Comput., 12(4), 2101–2109, 2016*.
- Pravda L, Berka K, Svobodová Vařeková R, Sehnal D, Banáš P, Laskowski RA, Koča J, Otyepka M: Anatomy of enzyme channels. *BMC Bioinf., 15(1), 379, 2014*.

OLEonline

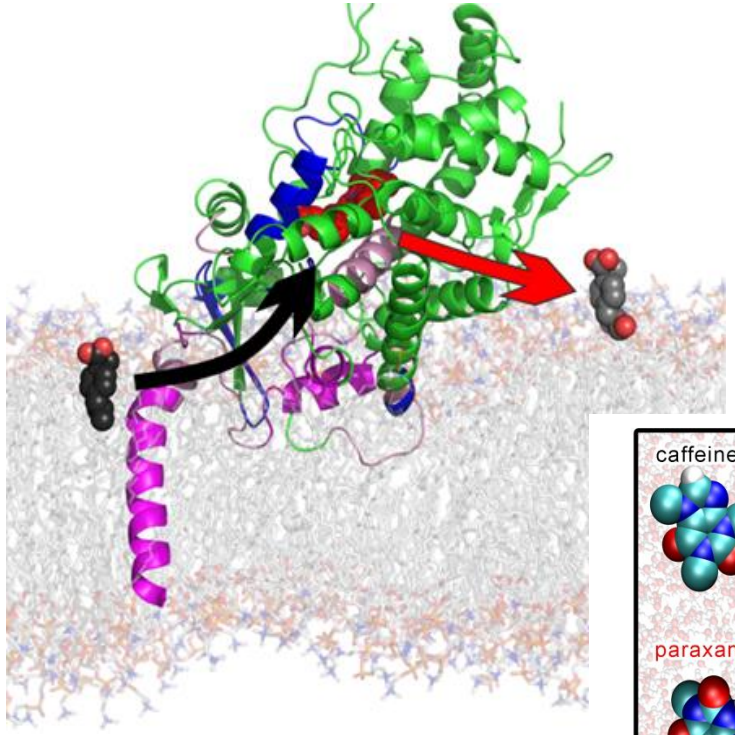
https://mole.upol.cz

- Detection and analysis of channels, tunnels and pores



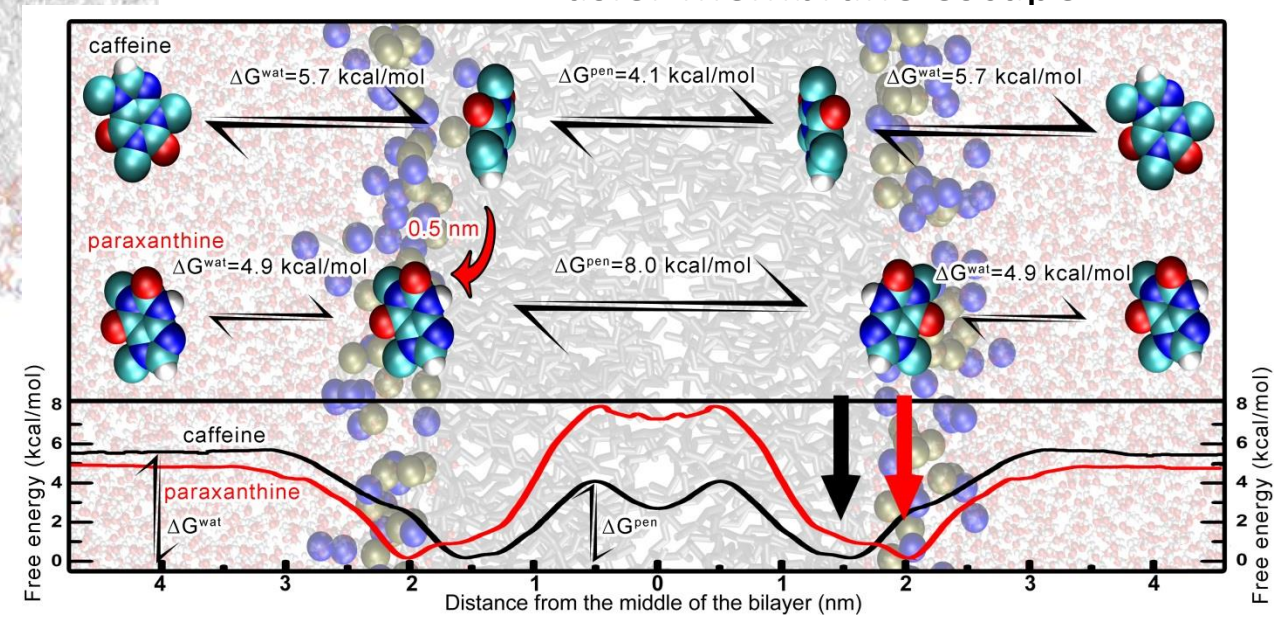
Pravda L, et al *Nucl. Acids Res,* 2018, doi:10.1093/nar/gky309

Application: Drugs with Cytochromes P450



- Drugs – mostly aromates
  - Deeper in the membrane
  - Higher lipids afinity
- Metabolites
  - Easier membrane escape

- Drugs
  - Access tunnel
- Metabolites
  - Eggress tunnel



Paloncýová M, **Berka K, Otyepka M**: *J. Phys. Chem. B* **2013**, *117*, 2403–10.
Berka K, **Paloncyova M**, Anzenbacher P, **Otyepka* M**: *J. Phys. Chem. B*, **2013,** 117(39), 11556-11564.
Berka K, Hendrychová T, Anzenbacher P, **Otyepka M**: *J. Phys. Chem. A* **2011**, *115*, 11248–11255.

X-RAY

Why X-Ray?

Elmag radiation interacts with objects of similar size with their wavelength (λ)

- visible light: λ = 350-700 nm and this is limit of optical microscopy

- RTG: CuKα: λ = 1,54 Å.

Synchrotron: λ = 0,5 Å – 2,5 Å.

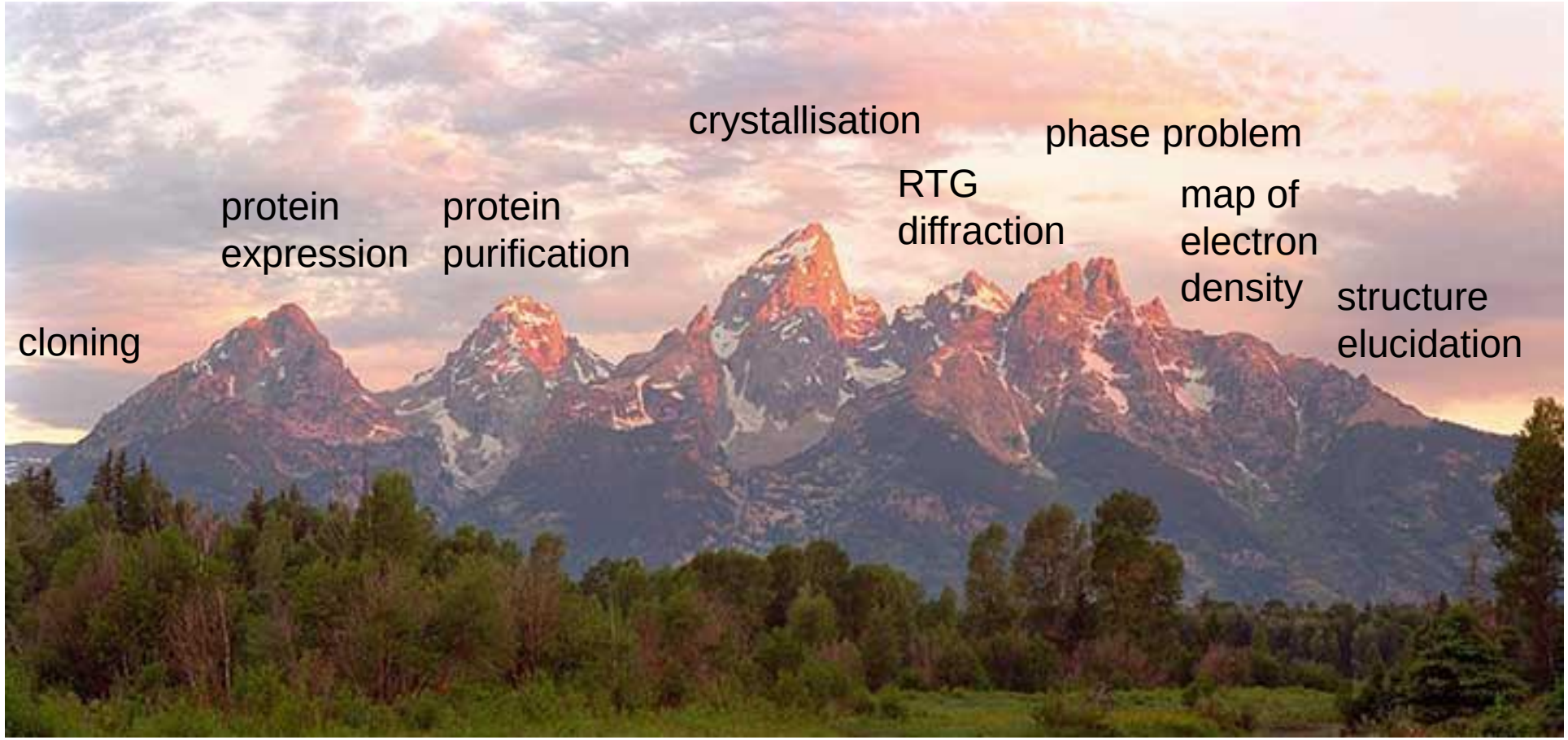
atom-atom distances: C-C = 1,54 Å,

C=C = 1,23 Å

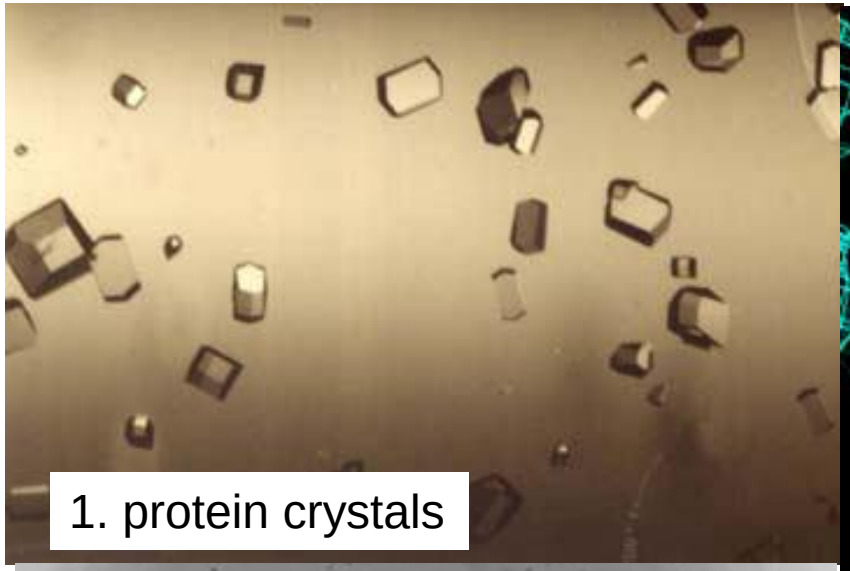
1 Å (Ångström) = 0.1 nm C-N = 1,45 Å

1 Å = 10^{-10} m N-(H).....O = 2,8 Å

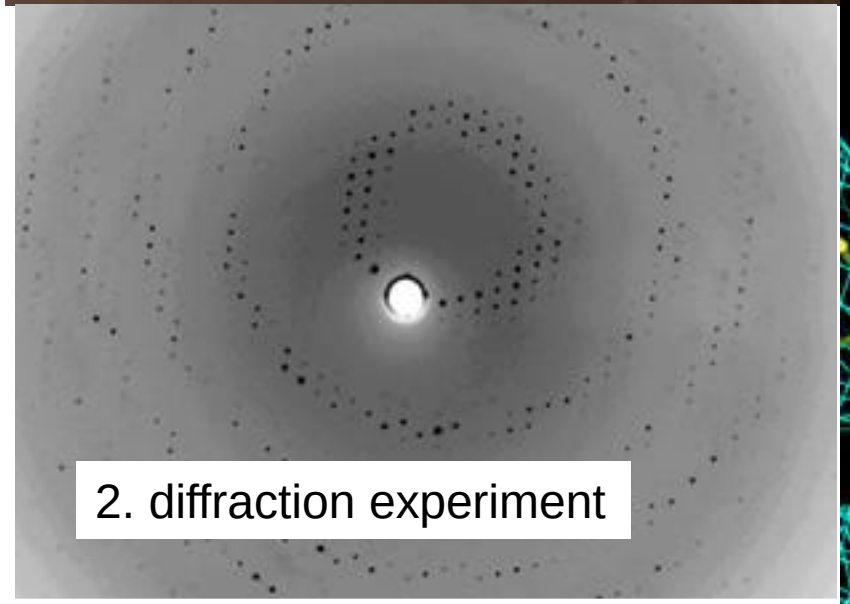
X-ray crystallography



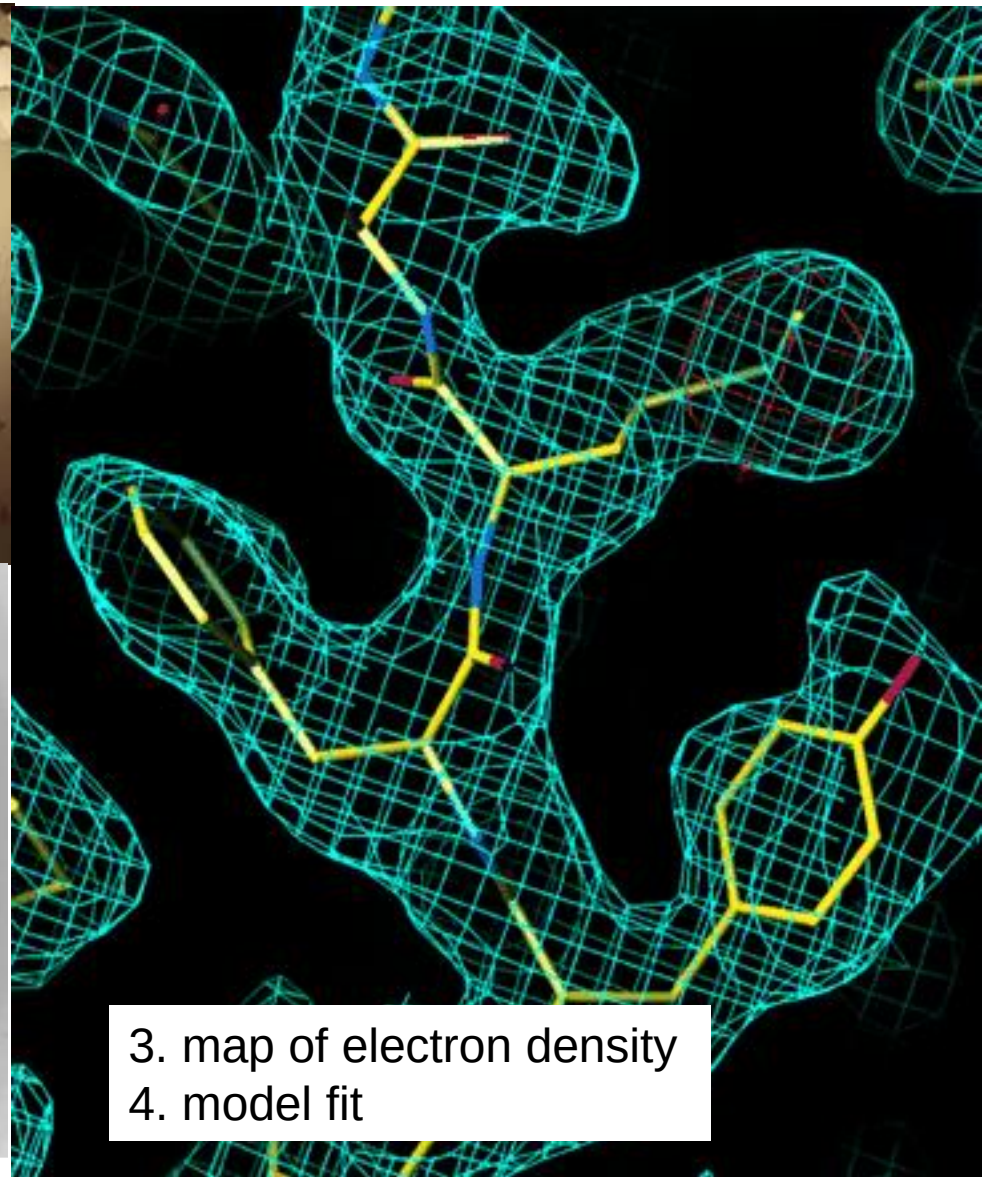
X-ray crystallography



1. protein crystals

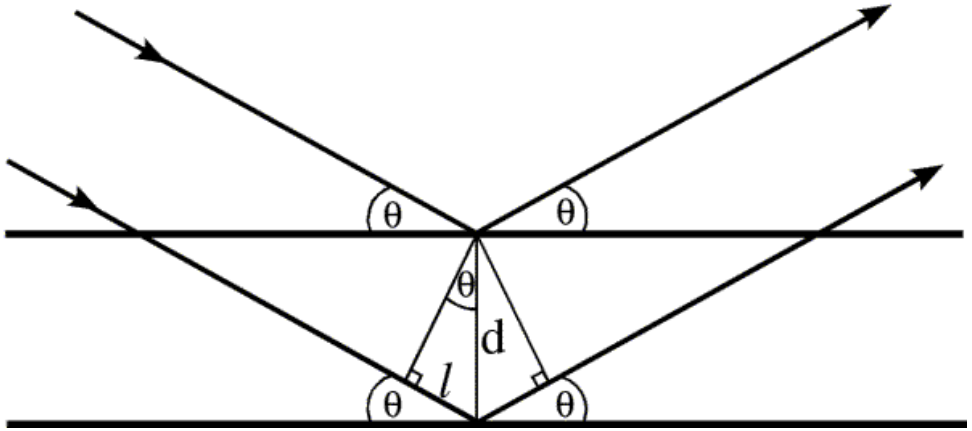


2. diffraction experiment



3. map of electron density
4. model fit

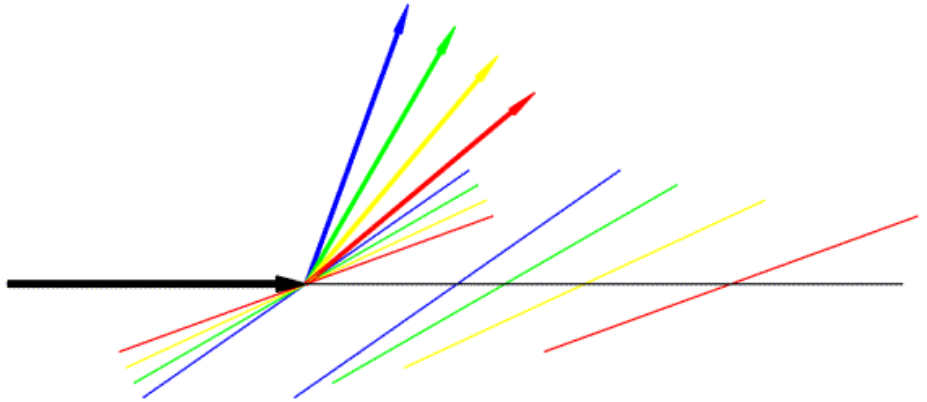
Diffraction Principle



Bragg's law

$n.\lambda = 2d.\sin\theta$

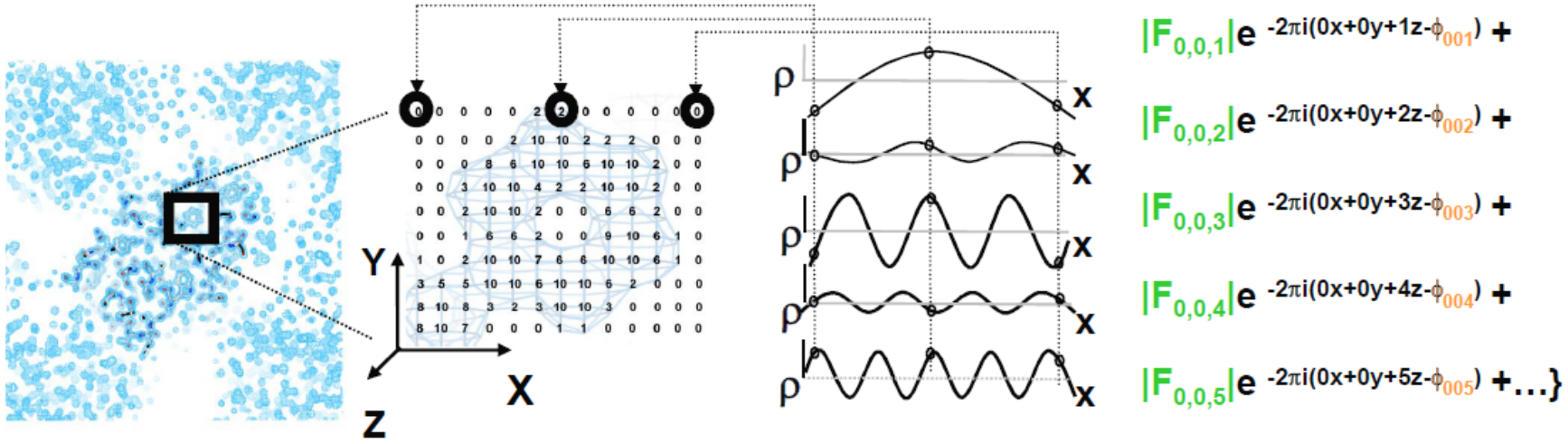
(W. H. Bragg & W. L. Bragg, 1912)



Diffracted radiation - sets of planes, parallel planes get boost in signal

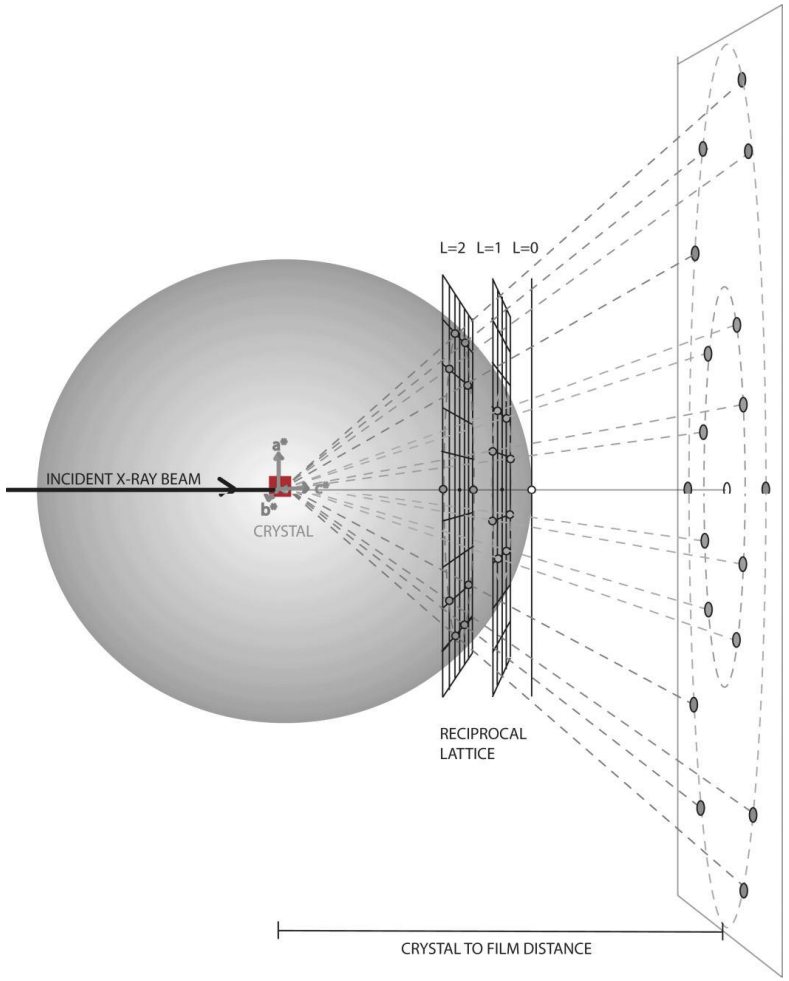
Calculation of Electron Density Map

Goal: use amplitudes and phases of thousands of diffractions F_{hkl} to calculate electron density map $\rho(x,y,z)$



$$\rho(x,y,z)=1/v\sum_h\sum_k\sum_l |F_{hkl}|e^{-2\pi i(hx+ky+lz-\phi_{hkl})}$$

Phase Problem



- Amplitudes and phases F_{hkl} are encoded in diffracted ray beams
- Amplitude $|F_{h,k,l}|$ is square root of intensity of diffracted beam.
- Φ_{hkl} is phase of diffracted wave. It cannot be directly measured – Phase problem.

Phase Problem

John C. Kendrew

F_1, Φ_1

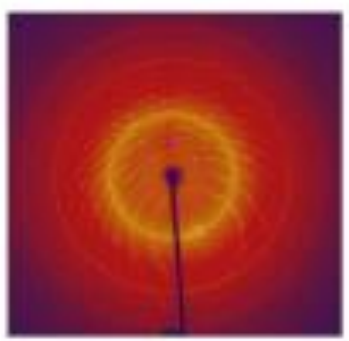
Max Perutz

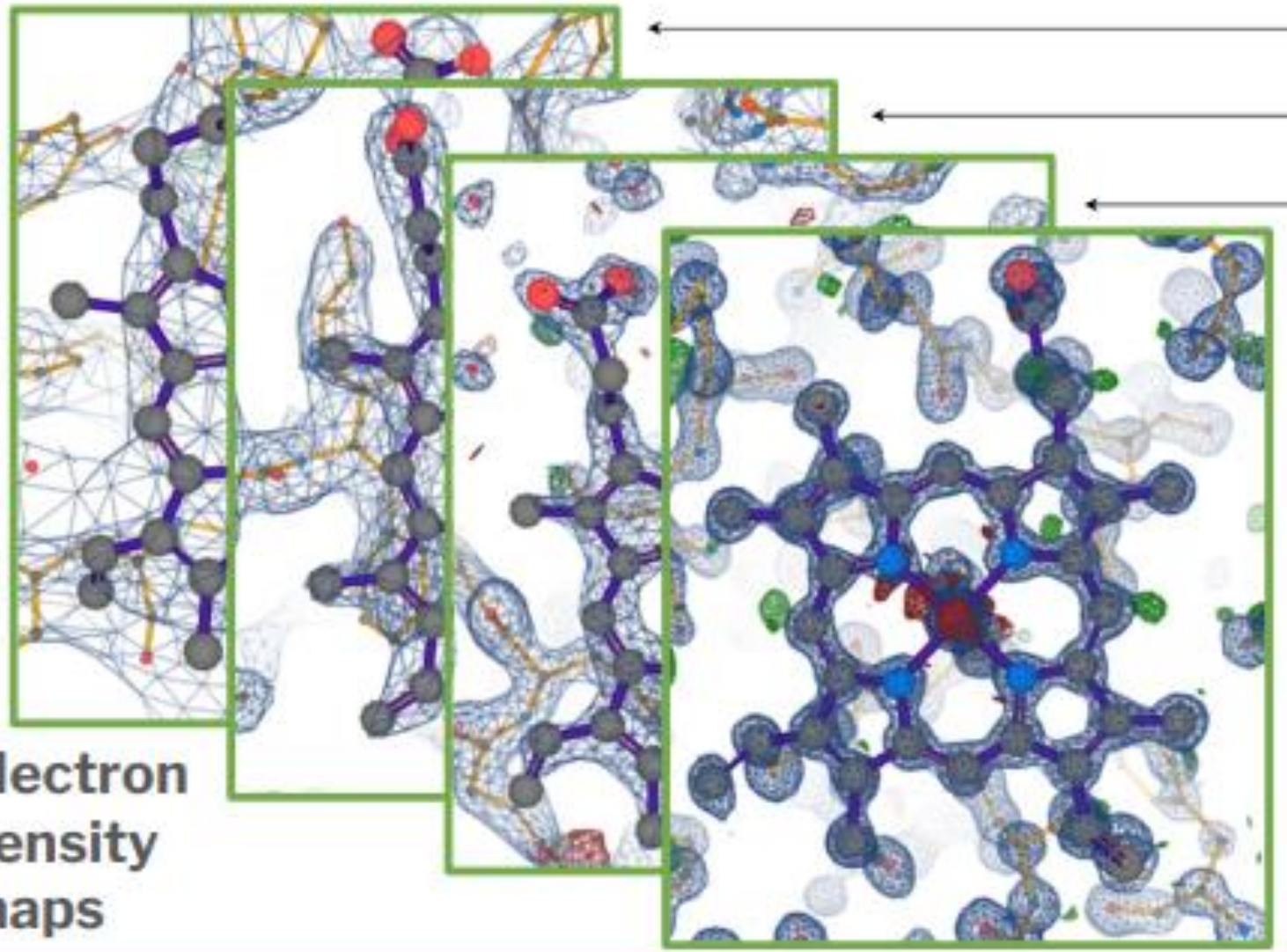
F_2, Φ_2

F_1, Φ_2

F_2, Φ_1

Maps are the real data - X-ray crystallography density maps





Electron density maps